EPA/600/X-08/006
December 2008

Method for Assessing the Toxicity of Sediment-Associated Contaminants with the Bivalve, *Mulinia lateralis*

Prepared by

M.C. Pelletier, USEPA, Narragansett, RI; M. Huber, SAIC, Narragansett, RI;
G.E. Morrison, USEPA, Narragansett, RI; R.M. Burgess, USEPA, Narragansett, RI;
G.M. Cripe, USEPA, Gulf Breeze, FL; and
P.A. Crocker, USEPA Region 6, Houston, TX

Office of Research and Development
U.S. Environmental Protection Agency

Disclaimer

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Abstract

This manual describes a laboratory method for determining the short-term toxicity of contaminated whole-sediments using the bivalve, *Mulinia lateralis.* Test sediments may be collected from estuarine or marine environments or spiked with compounds in the laboratory. The toxicity test is conducted for 10 days in 250-mL glass chambers containing 50 mL of sediment and 150 mL of overlying water. Overlying water is not renewed, and test organisms are fed phytoplankton during the toxicity tests. The test is maintained at 22 ±2 °C under constant light (2000-4000 lux). The salinity of the overlying water is maintained at 30 ±2‰, although *M. lateralis* is tolerant of salinities from 7‰ to >34‰. Although 30 ±2‰ is the standard, the test salinity can be adapted to lower salinities as long as these fall within the tolerance range of the species (please see section 4.3.6 for effects on metal bioavailability). *M. lateralis* has a wide range of grain size tolerance allowing greater latitude in the choice of negative and control sediment. The end points in the toxicity test are survival and growth (expressed as weight). Procedures are described for use with sediments from oligohaline to fully marine environments.

Foreword

Sediment contamination is a widespread environmental problem that can potentially pose a threat to a variety of aquatic ecosystems. Sediment functions as a reservoir for common contaminants such as pesticides, herbicides, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons, and metals such as lead, mercury, and arsenic. In-place contaminated sediment can result in depauperate benthic communities, while disposal of contaminated dredge material can potentially exert adverse effects on both pelagic and benthic systems. Historically, assessment of sediment quality has been limited to chemical characterizations. The U.S. Environmental Protection Agency (USEPA) is developing methodologies to calculate chemical-specific sediment quality criteria for use in the Agency's regulatory programs. However, quantifying contaminant concentrations alone cannot always provide enough information to adequately evaluate potential adverse effects that arise from interactions among chemicals, or that result from time-dependent availability of sediment-associated contaminants to aquatic organisms. Because relationships between concentrations of contaminants in sediment and bioavailability are not fully understood, determination of contaminated sediment effects on aquatic organisms often requires the use of controlled toxicity and bioaccumulation tests.

As part of USEPA's Contaminated Sediment Management Strategy, all Agency programs have agreed to use the same methods to determine whether sediments exhibit toxicity due to contamination. More than 10 Federal statutes provide authority to many USEPA program offices to address the problem of contaminated sediment. For example, the sediment toxicity test in this manual is currently only being considered for use in Region 6’s Ocean Dumping Program. However, it is the hope of the authors that presenting the method in this format will increase its use in nonregulatory activities related to dredged material disposal, registration of pesticides and toxic substances, Superfund site assessment, and assessment and cleanup of hazardous waste treatment, storage, and disposal facilities. The use of uniform sediment toxicity testing procedures by USEPA programs is expected to increase data accuracy and precision, facilitate test replication, increase the comparative value of test results, and, ultimately, increase the efficiency of processes requiring sediment toxicity tests.

Contents

Acknowledgments

This document is a general purpose testing manual for estuarine and marine sediments. The approaches were originally described in, Burgess et al. (1994) and Burgess and Morrison (1994). The current edition reflects the consensus of the U.S. Environmental Protection Agency (USEPA)—Narragansett, Gulf Breeze, Region 4, and Region 6—the Environmental Monitoring and Management Council (EMMC), and the Biological Advisory Committee.

The principal authors of this document are M.C. Pelletier, USEPA, Narragansett, RI; M. Huber, SAIC, Narragansett, RI; G.E. Morrison, USEPA, Narragansett, RI; R.M. Burgess, USEPA, Narragansett, RI; G.M. Cripe, USEPA, Gulf Breeze, FL; and P.A. Crocker, USEPA Region 6, Houston, TX. Contributors to specific sections of this manual are listed alphabetically below.

Sections 1-9, 11: General Guidelines

G.T. Ankley, USEPA, Duluth, MN
R.M. Burgess, USEPA, Narragansett, RI
G.A. Burton, Wright State University, Dayton, OH
F.J. Dwyer, NBS, Columbia, MO
R.A. Hoke, SAIC, Hackensack, NJ
M. Huber, SAIC, Narragansett, RI
C.J. Ingersoll, NBS, Columbia, MO
G.E. Morrison, USEPA, Narragansett, RI
T.J. Norberg-King, USEPA, Duluth, MN
M.C. Pelletier, USEPA, Narragansett, RI
C.E. Schlekat, SAIC, Narragansett, RI
K.J. Scott, SAIC, Narragansett, RI

Section 10: Collection, Culture and Maintenance of Test Organisms

G.M. Cripe, USEPA, Gulf Breeze, FL

Section 12: Statistical Analysis

J. Heltshe, SAIC, Narragansett, RI
R.A. Hoke, SAIC, Hackensack, NJ
H. Lee, USEPA, Newport, OR
T.J. Norberg-King, USEPA, Duluth, MN
C.E. Schlekat, SAIC, Narragansett, RI

Section 13: Precision and Accuracy

G.A. Burton, Wright State University, Dayton, OH
R.M. Burgess, USEPA, Narragansett, RI

M. Huber, SAIC, Narragansett, RI
G.E. Morrison, USEPA, Narragansett, RI
T.J. Norberg-King, USEPA, Duluth, MN
M.C. Pelletier, USEPA, Narragansett, RI
C.E. Schlekat, SAIC, Narragansett, RI
K.J. Scott, SAIC, Narragansett, RI

Additionally, USEPA Region 6 provided support for the development of this manual.

Section 1

Introduction

1.1 Significance of Use

1.1.1 Sediment provides habitat for many estuarine and marine organisms and is a major repository for many of the more persistent chemicals that are introduced into surface waters. In the aquatic environment, most anthropogenic chemicals and waste materials including toxic organic and inorganic chemicals eventually accumulate in sediment. Mounting evidence exists of environmental degradation in areas where USEPA Water Quality Criteria (WQC) are not exceeded, yet organisms in or near sediments are adversely affected (Chapman 1989). The WQC were developed to protect organisms in the water column and were not directed toward protecting organisms in sediment. Concentrations of contaminants in sediment may be several orders of magnitude higher than in the overlying water; however, bulk sediment concentrations have not been strongly correlated to bioavailability (Burton 1991). Partitioning or sorption of a compound between water and sediment may depend on many factors including aqueous solubility, pH, redox, affinity for sediment organic carbon and dissolved organic carbon, grain size of the sediment, sediment mineral constituents (oxides of iron, manganese, and aluminum), and the quantity of acid volatile sulfides in sediment (Di Toro et al. 1990, Di Toro et al. 1991). Although certain chemicals are highly sorbed to sediment, these compounds may still be available to the biota. Contaminated sediments may be directly toxic to aquatic life or can be a source of contaminants for bioaccumulation in the food chain.

1.1.2 Assessments of sediment quality have commonly included sediment chemical analyses and surveys of benthic community structure. Determination of sediment contaminant concentrations alone offers little insight into predicting adverse biological effects because bioavailability may be limited by the intricate partitioning factors mentioned above. Likewise, benthic community surveys may be inadequate because they sometimes fail to discriminate between effects of contaminants and those that result from unrelated noncontaminant factors, including water quality fluctuations, physical parameters, and biotic interactions. In order to obtain a direct measure of sediment toxicity, laboratory tests have been developed in which surrogate organisms are exposed to sediments under controlled conditions. Sediment toxicity tests have evolved into effective tools providing direct, quantifiable evidence of biological consequences of sediment contamination that can only be inferred from chemical or benthic community analyses. USEPA is developing a national inventory of contaminated sediment sites. This inventory will be used to develop a biennial report to Congress on sediment quality in the United States required under the Water Resources Development Act of 1992. The use of consistent sediment testing methods will provide high quality data needed for the national inventory and for regulatory programs to prevent, remediate, and manage contaminated sediment (Southerland et al. 1992).

1.1.3 The objective of a sediment test is to determine whether contaminants in sediment are harmful to or are bioaccumulated by benthic organisms. The tests can be used to measure interactive toxic effects of complex contaminant mixtures in sediment. Furthermore, knowledge of specific pathways of interactions among sediments and test organisms is not necessary in order to conduct the tests (Kemp and Swartz 1988). Sediment tests can be used to (1) determine the relationship between toxic effects and bioavailability, (2) investigate interactions among contaminants, (3) compare the sensitivities of different organisms, (4) determine spatial and temporal distribution of contamination, (5) evaluate hazards of dredged material, (6)

measure toxicity as part of product licensing or safety testing or chemical approval, (7) rank areas for clean up, and (8) set cleanup goals and estimate the effectiveness of remediation or management practices.

1.1.4 Results of toxicity tests on sediments spiked at different concentrations of contaminants can be used to establish cause and effect relationships between chemicals and biological responses. Results of toxicity tests with test materials spiked into sediments at different concentrations may be reported in terms of an LC50 (median lethal concentration), an EC50 (median effect concentration), an IC50 (inhibition concentration), or as a NOEC (no observed effect concentration) or LOEC (lowest observed effect concentration). However, spiked sediment may not be representative of contaminated sediment in the field. For example, mixing time (Stemmer et al. 1990a) and aging (Word et al. 1987, Landrum 1989, Landrum and Faust 1992) of spiked sediment can affect responses.

1.1.5 Evaluating effect concentrations for chemicals in sediment requires knowledge of factors controlling their bioavailability. Similar concentrations of a chemical in units of mass of chemical per mass of sediment dry weight often exhibit a range in toxicity in different sediments (Di Toro et al. 1990, Di Toro et al. 1991). Effect concentrations of chemicals in sediment have been correlated to interstitial water concentrations, and effect concentrations in interstitial water are often similar to effect concentrations in water-only exposures. The bioavailability of non-ionic organic compounds in sediment is often inversely correlated with the organic carbon concentration. Whatever the route of exposure, these correlations of effect concentrations to interstitial water concentrations indicate that predicted or measured concentrations in interstitial water can be used to quantify the exposure concentration to an organism. Therefore, information on partitioning of chemicals between solid and liquid phases of sediment is useful for establishing effect concentrations (Di Toro et al. 1991).

1.1.6 Field surveys can be designed to provide either a qualitative reconnaissance of the distribution of sediment contamination or a quantitative statistical comparison of contamination among sites. Surveys of sediment toxicity are usually part of more comprehensive analyses of biological, chemical, geological, and hydrographic data. Statistical correlations may be improved and sampling costs may be reduced if subsamples are taken simultaneously for sediment tests, chemical analyses, and benthic community structure.

1.1.7 Table 1.1 lists several approaches the USEPA has considered for the assessment of sediment quality (USEPA 1992c). These approaches include (1) equilibrium partitioning, (2) tissue residues, (3) interstitial water toxicity, (4) whole-sediment toxicity and sediment-spiking tests, (5) benthic community structure, and (6) Sediment Quality Triad and Apparent Effects Thresholds (see Chapman 1989, USEPA 1989a, USEPA 1990a, USEPA 1990b, and USEPA 1992b for a critique of these methods). The sediment assessment approaches listed in Table 1.1 can be classified as numeric (e.g., equilibrium partitioning), descriptive (e.g., whole-sediment toxicity tests), or a combination of numeric and descriptive approaches (e.g., Apparent Effects Threshold; USEPA 1992c, Long et al. 1995). Numeric methods can be used to derive chemical-specific sediment quality criteria (SQC). Descriptive methods such as toxicity tests with field-collected sediment cannot be used alone to develop numerical SQC for individual chemicals. Although each approach can be used to make site-specific decisions, no one single approach can adequately address sediment quality. Overall, an integration of several methods using the weight of evidence is the most desirable approach for assessing the effects of contaminants associated with sediment (Long and Morgan 1990). Hazard evaluations integrating data from laboratory exposures, chemical analyses, and benthic community.

Table 1.1 SEDIMENT QUALITY ASSESSMENT PROCEDURES
(Modified from USEPA [1992c])

Method	Type			Approach
	Numeric	Descriptive	Combination	
Equilibrium Partitioning	*			A sediment quality value for a given contaminant is determined by calculating the sediment concentration of the contaminant that corresponds to an interstitial water concentration equivalent to the USEPA water quality criterion for the contaminant.
Tissue Residues	*			Safe sediment concentrations of specific chemicals are established by determining the sediment chemical concentration that results in acceptable tissue residues.
Interstitial Water Toxicity	*	*	*	Toxicity of interstitial water is quantified and identification evaluation procedures are applied to identify and quantify chemical components responsible for sediment toxicity.
Benthic Community Structure		*		Environmental degradation is measured by evaluating alterations in benthic community structure.

Table 1.1 SEDIMENT QUALITY ASSESSMENT PROCEDURES *(cont'd.)*
(Modified from USEPA [1992c])

Whole-Sediment Toxicity and Sediment Spiking	*	*	*	Test organisms are exposed to sediments that may contain known or unknown quantities of potentially toxic chemicals. At the end of a specified time period, the response of the test organisms is examined in relation to a specified endpoint. Dose-response relationships can be established by exposing test organisms to sediments that have been spiked with known amounts of chemicals or mixtures of chemicals.
Sediment Quality Triad	*	*	*	Sediment chemical contamination, sediment toxicity, and benthic community structure are measured on the same sediment sample. Correspondence between sediment chemistry, toxicity, and field effects is used to determine sediment concentrations that discriminate conditions of minimal, uncertain, and major biological effects.

Table 1.1 SEDIMENT QUALITY ASSESSMENT PROCEDURES *(cont'd.)*
(Modified from USEPA [1992c])

Apparent Effects Threshold	*	*	*	The sediment concentration of a contaminant above which statistically significant biological effects (e.g., sediment toxicity) are always expected. AET values are empirically derived from paired field data for sediment chemistry and a range of biological effects indicators.

assessments (the Sediment Quality Triad) provide strong complementary evidence of the degree of pollution-induced degradation in aquatic communities (Chapman et al. 1992, Burton 1991).

1.2 Program Applicability

1.2.1 The USEPA has authority under a variety of statutes to manage contaminated sediment. Until recently, the USEPA has not addressed sediment quality except in relation to disposal of material removed during navigational dredging (Table 1.2). Southerland et al. (1992) outlined four goals of a USEPA management strategy for contaminated sediments: (1) in-place sediment should be protected from contamination to ensure beneficial uses of surface waters, (2) protection of in-place sediment should be achieved through pollution prevention and source control, (3) in-place remediation should be limited to locations where natural recovery will not occur in an acceptable period of time, and (4) consistent methods should be used to trigger regulatory decisions.

1.2.2 The Clean Water Act (CWA) is the single most important law dealing with environmental quality of surface waters in the United States. The goal of the CWA is to restore and maintain physical, chemical, and biological integrity of the nation's waters (Southerland et al. 1992). Federal and state monitoring programs traditionally have focused on evaluating water column problems caused by point-source dischargers. During the next few years, the USEPA will be developing a national inventory of contaminated sediment sites. This inventory will be used to develop a biennial report to Congress on sediment quality in the United States required under the Water Resources Development Act of 1992. The use of consistent sediment testing methods will provide high quality data needed for the national inventory and for regulatory program to prevent, remediate, and manage contaminated sediment (Southerland et al. 1992).

1.2.3 The Office of Water (OW), the Office of Pesticide Programs (OPP), the Office of Pollution Prevention and Toxic Substances (OPPTS), the Office of Solid Waste (OSW), and the Office of Emergency and Remedial Response (OERR) are all committed to the principle of consistent tiered testing outlined in the Agency-wide Contaminated Sediment Strategy (Southerland et al.

Table 1.2 STATUTORY NEEDS FOR SEDIMENT QUALITY ASSESSMENT
(Modified from Dickson et al. 1984 and Southerland et al. 1992)

Law [a]	Area of Need
CERCLA	• Assess need for remedial action with contaminated sediments; assess degree of cleanup required, disposition of sediments
CWA	• NPDES permitting, especially under Best Available Technology (BAT) in water-quality-limited water • Section 403(c) criteria for ocean discharges; mandatory additional requirements to protect marine environment • Section 301(g) waivers for publicly owned treatment works (POTWS) discharging to marine waters • Section 404 permits for dredge and fill activities (administered by the Corps of Engineers)
FIFRA	• Review uses of new and existing chemicals • Pesticide labeling and registration
MPRSA	• Permits for ocean dumping
NEPA	• Preparation of environmental impact statements for projects with surface water discharges
TSCA	• Section 5: Premanufacture notice reviews for new chemicals • Sections 4, 5, and 6: Reviews for existing chemicals
RCRA	• Assess suitability (and permit) on-land disposal or beneficial use of contaminated sediments considered “hazardous”

[a] CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act (“Superfund”)
CWA	Clean Water Act
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
MPRSA	Marine Protection, Research and Sanctuary Act
NEPA	National Environmental Policy Act
TSCA	Toxic Substances Control Act
RCRA	Resource Conservation and Recovery Act

1992). Agency-wide consistent testing is desirable because all USEPA programs will use similar methods to evaluate whether a sediment poses an ecological or human health risk, and comparable data would be produced. It will also provide the basis for uniform cross-program

decision making within the USEPA. Each program will, however retain the flexibility of deciding whether identified risks would trigger regulatory actions.

1.2.4 Tiered testing should include a hierarchy of tests with the tests in each successive tier becoming progressively more rigorous, complex, and costly (Southerland et al. 1992). Guidance needs to be developed to explain how information within each tier would trigger regulatory action. The guidance could be program specific, describing decisions based on a weight of evidence approach, a pass-fail approach, or comparison to a reference site depending on statutory and regulatory requirements. There are now two approaches for tiered testing used by USEPA:

(1) the Office of Water-U.S. Army Corps of Engineers dredged material testing framework and

(2) the OPP ecological risk assessment tiered testing framework.

Tier 1 of the dredged material testing framework consists of a review of existing chemical and biological data or an inventory of nearby sources. In Tier 2, chemical data are compared to water and sediment quality criteria. Tier 3 evaluations consist of acute toxicity and bioaccumulation testing, and a comparison of the results to a reference area. Tier 4 studies consist of site-specific field studies. The OPP testing framework consists of acute toxicity testing in Tier 1, followed by chronic (early life stage) toxicity testing in Tier 2 and further chronic toxicity testing (full life cycle) in Tier 3. A tiered testing framework has not yet been chosen for Agency-wide use, but some of the components have been identified to be standardized. These components are toxicity tests, bioaccumulation tests, chemical criteria, and other measurements that may have ecological significance including benthic community structure evaluation, colonization rate, and *in situ* sediment testing within a mesocosm (USEPA 1992a).

1.3 Role of Bivalve, *Mulinia lateralis,* Sediment Toxicity Test

1.3.1 One of the roles of the test described in this document is to eventually serve as a supplement to the existing suite of marine sediment toxicity tests. These animals used in sediment testing have been limited to crustaceans, and their habitats are very similar to the mollusc, *Mulinia lateralis* (Table 1.3). Other roles include uses as a scientific (e.g., bioavailability studies) and a nonregulatory site assessment tool. The method (Burgess and Morrison 1994) has not undergone the extensive review and validation procedures required by USEPA regions and program offices. Before this method can be considered or used as a valid test it must undergo these reviews and validations.

1.3.2 The USEPA Region 6 supported a method evaluation and validation exercise in which laboratories around the country used the method with reference toxicants. This process served several objectives (i.e., (1) familiarize scientific community with method, (2) identify method strengths and weaknesses, and (3) measure interlaboratory precision) and the results are being evaluated to determine the reliability of the bivalve, *Mulinia lateralis,* sediment toxicity test. Until completion of that process, the method could be considered for nonregulatory applications.

1.4 Scope and Application

1.4.1 Procedures are described for testing the euryhaline bivalve, *Mulinia lateralis* in the laboratory to evaluate the toxicity of contaminants associated with whole sediments. Sediments

Table 1.3 COMPARISON OF HABITAT CHARACTERISTICS AND OTHER LIFE HISTORY PARAMETERS OF FOUR ESTUARINE AND MARINE AMPHIPOD SPECIES TO *MULINIA LATERALIS* USED IN SEDIMENT TOXICITY TESTS

Characteristic	*Ampelisca abdita*	*Eohaustorius estuarius*	*Leptocheirus plumulosus*	*Rhepoxynius abronius*	*Mulinia lateralis*
Substrate Relation	Tube dwelling, closed [a]	Free burrowing [e]	Tube dwelling, open [a]	Free burrowing [j]	Free burrowing [m]
Zoogeography	Atlantic-Gulf [a,b,c]	Pacific [e,f]	Atlantic [a]	Pacific [j]	Atlantic-Gulf of Mexico [m]
Habitat	Poly-upper mesohaline [a]	Oligo-mesohaline [e,f]	Oligo-mesohaline [a]	Polyhaline [j,k]	Euryhaline [m]
Life Cycle	60 to 120 days [d]	Annual [e]	30 to 40 days [g]	Annual [l]	60 to 120 days [m]
Availability	Field-culture [d]	Field [e]	Field-culture [g,h,i]	Field [k]	Field-culture [m]
Ecological Importance	High	High	High	High [k]	High [m]

a Bousfield 1973
b Nichols and Thompson 1985
c Hopkins 1986
d Scott and Redmond 1989
e DeWitt et al. 1989
f Environment Canada 1992
g DeWitt et al. 1992
h Schlekat et al. 1992
i McGee et al. 1993
j Barnard and Barnard 1982
k Swartz et al. 1985
l Kemp et al. 1985
m Calabrese 1970

---

may be collected from the field or spiked with toxicants in the laboratory. *Mulinia lateralis* is found along the Atlantic coast from Prince Edward Island, Canada to northeastern Mexico and the West Indies. The toxicity test is conducted for 10-days in 250 mL glass chambers containing 50 mL of sediment and 150 mL of overlying seawater. Exposure is static (i.e., water is not renewed), and the animals are fed phytoplankton over the 10-d exposure period. The end points in the toxicity test are survival and growth. Procedures are described for use with sediments with pore water salinity ranging from >0‰ to fully marine.

1.4.2 This manual represents a new area of marine sediment toxicity testing research; that is, use of a bivalve species with both acute and sublethal endpoints. Other research and methods development are now in progress to (1) develop standard sediment bioaccumulation tests (i.e., 28-day exposures with the bivalve *Macoma nasuta* and the polychaete *Nereis virens)* (Lee et al. 1989), (2) develop standard chronic sediment toxicity tests (e.g., 28-day exposures with

Leptocheirus plumulosus), (3) refine sediment spiking procedures, (4) refine sediment dilution procedures, (5) refine sediment Toxicity Identification Evaluation (TIE) procedures, and (6) produce additional data on confirmation of responses in laboratory tests with natural populations of benthic organisms. This information will be described in future editions of the manual.

1.4.3 Procedures described in this manual are based on the following documents: Burgess et al. (1994), Burgess and Morrison (1994), and Cripe (2006). While standard procedures are described in the manual, further investigation of certain issues could aid in the interpretation of test results. Some of these issues include the effect of shipping on organism sensitivity, the effect of testing at different salinities on organism response, additional performance criteria for organism health, and confirmation of responses in laboratory tests with natural benthic populations.

1.4.4 Altering the procedures described in this manual may alter bioavailability and produce results that are not directly comparable with results using the accepted procedure. Comparison of results obtained using a modified version of this procedure might provide useful information concerning new concepts and procedures for conducting sediment tests with aquatic organisms. If tests are conducted with methods different from those described in this manual, additional tests are required to determine comparability of results.

1.4.5 Where States have developed culturing and testing methods for indigenous species other than the one recommended in the manual, data comparing the sensitivity of the substitute species and the recommended species must be obtained with sediments or reference toxicants, to ensure that the species selected are at least as sensitive and appropriate as the recommended species.

1.4.6 Selection of Toxicity Testing Organism

1.4.6.1 The choice of a test organism has a major influence on the relevance, success, and interpretation of a test. Test organism selection should be based on both environmental relevance and practical concerns (DeWitt et al. 1989; Swartz 1989). Ideally, a test organism should (1) have a toxicological database demonstrating relative sensitivity to a range of contaminants of interest in sediment, (2) have a database for interlaboratory comparisons of procedures (e.g., round-robin studies), (3) be in direct contact with sediment, (4) be readily available from culture or through field collection, (5) be easily maintained in the laboratory, (6) be easily identified, (7) be ecologically or economically important, (8) have a broad geographical distribution, be indigenous (either present or historical) to the site being evaluated, or have a niche similar to organisms of concern (e.g., similar feeding guild or behavior to the indigenous organisms), (9) be tolerant of a broad range of sediment physico-chemical characteristics (e.g., grain size), and (10) be compatible with selected exposure methods and end points. Methods utilizing selected organisms should also be (11) peer reviewed (e.g., journal articles, ASTM guides) and (12) confirmed with responses of natural populations of benthic organisms.

1.4.6.2 The bivalve, *M. lateralis,* has several unique qualities which support its use for toxicity testing. First, juvenile bivalves have intimate contact with sediments and are relatively sessile and unable to migrate from contaminated areas. Further, when burrowing they resuspend sediment that, subsequently, is ingested during filter feeding. Second, although *M. lateralis* is relatively insensitive to organic toxicants, it is very sensitive to metals. This insensitivity to organic toxicants suggests the bivalve may also serve well as a bioaccumulation organism. Third, larval and juvenile bivalves have demonstrated extreme sensitivity to photoinduced PAHs (Pelletier 1997). Finally, in a comparison with 62 contaminated sediments, juvenile *M. lateralis*

demonstrated comparable acute toxicity to that of the amphipods *A. abdita* and *E. estuarius* (Burgess and Morrison 1994). However, bivalve growth was often a more sensitive endpoint than amphipod mortality (Burgess and Morrison 1994).

1.4.6.3 An important consideration in the selection of specific species for standardization of a test method development is the existence of information concerning relative sensitivity of the organisms both to single chemicals and complex mixtures. Several bivalve genera were among the test species used to generate saltwater Water Quality Criteria for 26 chemicals. Although *Mulinia* was not among the genera in the database, other work has indicated that *M. lateralis* shows comparable sensitivity to commercially exploited bivalve species such as mussels, oysters, and hard clams. The bivalve data were compared to other species data for (1) all chemicals, (2) metals, and (3) organics. Average percentile rank for bivalves was 58%, indicating median sensitivity. There was a slight increase in sensitivity for metals (49%) although bivalves showed extreme sensitivity to certain metals. A slight decrease in sensitivity was seen for organics (73%) although bivalves showed increased sensitivity for certain organics. Overall, these data indicate that bivalves are not uniquely sensitive to all compounds relative to other species. Comparisons of contaminated sediments using amphipods and juvenile *M. lateralis* indicate that these species show complementary sensitivity.

1.4.6.4 Data from 18 community tests (colonization tests) with pesticides, formulated products, organic chemicals and drilling muds indicated that the phylum Mollusca was frequently the most sensitive. In these tests the early life stages of settling organisms were exposed (Hansen 1984). Later community tests performed with 1,2,4-trichlorobenzene showed that molluscs were more sensitive than arthropods. This sensitivity was observed for both settled animals (juveniles) exposed to spiked overlying water and larvae and settling stages exposed to spiked sand. Early life stages were shown to be twice as sensitive as the settled stages (Tagatz et al. 1985).

1.4.6.5 Limited comparative data is available for water-only exposures of *M. lateralis* and other species in single-chemical tests. Studies that do exist generally show that no one species is consistently the most sensitive.

1.4.6.5.1 *M. lateralis* is relatively sensitive to ammonia. The water-only LC50 (pH 7.79, SD: 0.20) for total ammonia is 21.7 mg/L (overlying seawater), and for unionized ammonia the water-only LC50 is 0.6 mg/L. The water-only EC50 (growth) (pH 7.79, SD: 0.20) for total ammonia is 11.0 mg/L, and for unionized ammonia, the water-only EC50 is 0.3 mg/L. For survival, the water-only NOEC is 8.2 mg/L for total ammonia and 0.2 mg/L for unionized ammonia. The water-only NOEC for growth is less than 2.4 mg/L and for unionized ammonia, is less than 0.1 mg/L). The amphipod *A. abdita* has a total ammonia LC50 of 43.9 at pH 7.7, and the unionized LC50 is 0.8. At pH 8.4, the total ammonia LC50 for the amphipod was 13.6 mg/L, and the unionized LC50 was 1.1 mg/L.

1.4.6.5.2 The sensitivity of *M. lateralis* to ammonia is of concern as many marine sediments have elevated pore-water levels, especially those impacted by anthropogenic activity (i.e., eutrophication). For development of this document, several sediments with pore water ammonia level ranging from 4-73 mg/L were placed into the *M. lateralis* toxicity test exposure chambers (described in Chapter 11). Overlying water was added to the sediment and, 24 h later, the overlying water ammonia levels were 0.7-11 mg/L, approximately 75% to 85% less than porewater ammonia. The method calls for an initial “flushing” on day 0. Ammonia levels 24-h after performing this procedure were 0.3-5 mg/L, approximately 94% to 120% less than initial overlying water levels (Huber et al. 1997). By the end of the test (day 10) overlying water ammonia levels were below detection limit (~0.1 mg/L). As *M. lateralis* is exposed to toxicants

predominantly via the overlying waters, moderate levels of pore water ammonia should not be a concern when the flushing procedure is followed.

1.4.6.5.3 Cadmium and copper have been used as common reference toxicants. Seven-day water-only renewal tests with both toxicants were conducted with *Mulinia lateralis* juveniles to determine the LC50 (survival) and EC50 (growth). Similarly, 10-day water-only static-renewal tests were conducted to determine the LC50 (survival) for *Ampelisca abdita.* For *M. lateralis,* the cadmium LC50 is 1849 µg/L (1795-1947 µg/L, 95% C.I.) and the EC50 is 897 µg/L (819-946 µg/L, 95% C.I.)(M. Pelletier 1997). The *A. abdita* cadmium LC50 is 36 µg/L (Berry et al. 1996). For *M. lateralis,* the copper LC50 is 52 µg/L (27-70 µg/L, 95% C.I.) and the EC50 for copper is 27 µg/L (23-27 µg/L, 95% C.I.)(M. Pelletier 1997). The *A. abdita* copper LC50 is 20.5 µg/L (Berry et al. 1996).

1.4.6.6 Relative species sensitivity frequently varies among contaminants; consequently, a battery of tests including organisms representing different phylogenetic groups may be needed to assess sediment quality (Craig 1984, Williams et al. 1986, Long et al. 1990, Ingersoll et al. 1990, Burton and Ingersoll 1994). For example, Reish (1988) reported the relative toxicity of six metals (arsenic, cadmium, chromium, copper, mercury, and zinc) to crustaceans, polychaetes, pelecypods, and fishes and concluded that no one species or group of test organisms was the most sensitive to all of the metals.

1.4.6.7 The sensitivity of an organism is related to route of exposure and biochemical response to contaminants. Sediment-dwelling organisms can receive a dose from three primary sources: interstitial water, sediment particles, and overlying water. Food type, feeding rate, assimilation efficiency, and clearance rate will control the dose of contaminants from sediment. Benthic invertebrates often selectively consume different particle sizes (Harkey et al. 1994) or particles with higher organic carbon concentrations that may have higher contaminant concentrations. Grazers and other collector-gatherers that feed on particles and detritus may receive most of their body burden directly from materials attached to sediment or from actual sediment ingestion. In some amphipods (Landrum 1989) and clams (Boese et al. 1990) uptake through the gut can exceed uptake across the gills for certain hydrophobic compounds. Organisms in direct contact with sediment may also accumulate contaminants by direct adsorption to the body wall or by absorption through the integument (Knezovich et al. 1987).

1.4.6.8 Despite the potential complexities in estimating the dose that an animal receives from sediment, the toxicity and bioaccumulation of many contaminants in sediment such as Kepone, fluoranthene, organochlorines, and metals have been correlated with either the concentration of these chemicals in interstitial water or in the case of non-ionic organic chemicals, concentrations on an organic-carbon basis (Di Toro et al. 1990, Di Toro et al. 1991). The relative importance of whole sediment and interstitial water routes of exposure depends on the test organism and the specific contaminant (Knezovich et al. 1987). Because benthic communities contain a diversity of organisms, many combinations of exposure routes may be important. Therefore, behavior and feeding habits of a test organism can influence its ability to accumulate contaminants from sediment and should be considered when selecting test organisms for sediment testing.

1.5 Performance Criteria and Test Acceptability

1.5.1 USEPA's Environmental Monitoring Management Council (EMMC) recommended the use of performance-based methods in developing chemical analytical standards (Williams 1993). Performance-based methods were defined by EMMC as a monitoring approach which permits

the use of appropriate methods that meet pre-established demonstrated performance standards (section 9.2).

1.5.2 The key consideration for methods used to obtain test organisms, whether they are field-collected or obtained from culture, is having healthy organisms of known quality. A performance-based criteria approach was selected as the preferred method through which individual laboratories should evaluate culture methods or the quality of field-collected organisms rather than by control-based criteria. This method was chosen to allow each laboratory to optimize culture methods, determine the quality of field-collected organisms, and minimize effects of test organism health on the reliability and comparability of test results. See Table 1.3 for a listing of performance criteria used to assess the quality of cultured *M. lateralis* and to determine the acceptability of 10-day sediment toxicity tests.

Table 1.3 RATING OF SELECTION CRITERIA FOR *MULINIA LATERALIS*
(A "+" "–" indicates a positive or negative attribute, respectively.)

Criterion	*Mulinia lateralis*
Relative sensitivity toxicity database	+
Round-robin studies conducted	-
Contact with sediment	+
Laboratory culture	+
Maintain in laboratory	+
Taxonomic identification	+
Ecological importance	+
Geographical distribution [a]	ATL, GM
Sediment physico-chemical tolerance	+
Field-validated	-
Peer-reviewed	+
End points monitored	Survival, Growth

[a] ATL = Atlantic Coast, GM = Gulf of Mexico

Section 2

Summary of Method

2.1 Method Description and Experimental Design

2.1.1 Method Description

2.1.1.1 This manual describes a laboratory method for determining the short-term toxicity of contaminated whole-sediments using the bivalve *Mulinia lateralis.* Test sediments may be collected from estuarine or marine environments or spiked with toxicants in the laboratory. The toxicity test is conducted for 10 days in 250-mL glass chambers containing 50 mL of sediment and 150 mL of overlying water without aeration. The number of treatments and replicates will depend on the objective of the study (see section 2.1.2.2). Overlying water is not renewed, and test organisms are fed phytoplankton during the toxicity tests. The test is maintained at 22 ±2 °C under constant light (2000- 4000 lux). The salinity of the overlying water is maintained at 30 ±2‰. Although 30 ±2‰ is the standard, the test salinity can be adapted to lower salinities as long as these fall within the tolerance range of the species (please see section 4.3.6 for effects on metal bioavailability). *M. lateralis* has a wide range of grain size tolerance allowing greater latitude in the choice of negative and positive control sediment. End points in the toxicity test are survival and growth. Procedures are described for use with sediments from oligohaline to fully marine environments.

2.1.2 Experimental Design. The following section is a general summary of experimental design (see section 12 for additional detail).

2.1.2.1 Control and Reference Sediment.

2.1.2.1.1 Sediment tests include a control sediment (sometimes called a negative control). A control sediment is a sediment that is essentially free of contaminants and is used routinely to assess the acceptability of a test and is not necessarily collected near the site of concern. Any contaminants in control sediment are thought to originate from the global spread of pollutants and do not reflect any substantial input from local or nonpoint sources (Lee et al. 1992). A control sediment provides a measure of test acceptability, evidence of test organism health, and a basis for interpreting data obtained from the test sediments. A reference sediment is collected near an area of concern and is used to assess sediment conditions exclusive of material(s) of interest. Testing a reference sediment provides a site-specific basis for evaluating toxicity.

2.1.2.1.2 Natural geomorphological and physico-chemical characteristics such as sediment texture may influence the response of test organisms (DeWitt et al. 1988). The physico-chemical characteristics of test sediment must be within the tolerance limits of the test organism. *M. lateralis* has a wide range of tolerance of sediment grain size. See Table 11.4 for tolerance limits of *M. lateralis* for physico-chemical characteristics. If the physico-chemical characteristic(s) of a test sediment exceed the tolerance limits, it may be desirable to include a control sediment that encompasses those characteristics. The effects of some sediment characteristics on the results of sediment tests may be able to be addressed with regression equations (DeWitt et al. 1988, Ankley et al. 1994).

2.1.2.2 The experimental design depends on the purpose of the study. Variables that need to be considered include the number and type of control sediments, the number of treatments and

replicates, and water quality characteristics. For instance, the purpose of the study might be to determine a specific end point such as an LC50 and may include a control sediment (a positive control), and several concentrations of sediment spiked with a chemical. A useful summary of field sampling design is presented by Green (1979). See section 12 for additional guidance on experimental design and statistics.

2.1.2.3 If the purpose of the study is to conduct a reconnaissance field survey to identify contaminated sites for further investigation, the experimental design might include only one sample from each site to allow for maximum spatial coverage. The lack of replication at a site usually precludes statistical comparisons (e.g., ANOVA), but these surveys can be used to identify contaminated sites for further study or may be evaluated using regression techniques (Sokal and Rohlf 1981, Steel and Torrie 1980).

2.1.2.4 In other instances, the purpose of the study might be to conduct a quantitative sediment survey to determine statistically significant differences between effects among control and test sediments from several sites. The number of replicates/site should be based on the need for sensitivity or power (section 12). In a quantitative survey, replicates (separate samples from different grabs collected at the same site) would need to be taken at each site. Chemical and physical characteristics of each of these grabs would be required for sediment testing. Separate subsamples might be used to determine within-sample variability or for comparisons of test procedures (e.g., comparative sensitivity among test organisms), but these subsamples cannot be considered to be true field replicates for statistical comparisons among sites (ASTM 1993a).

2.1.2.5 Sediments often exhibit high spatial and temporal variability (Stemmer et al. 1990a). Therefore, replicate samples may need to be collected to determine variance in sediment characteristics. Sediment should be collected with as little disruption as possible; however, subsampling, compositing, or homogenization of sediment samples may be necessary for some experimental designs.

2.1.2.6 Site locations might be distributed along a known pollution gradient, in relation to the boundary of a disposal site, or at sites identified as being contaminated in a reconnaissance survey. Both spatial and temporal comparisons can be made. In predredging studies, a sampling design can be prepared to assess the contamination of samples representative of the project area to be dredged. Such a design should include subsampling cores taken to the project depth.

2.1.2.7 The primary focus of the physical and experimental test design, and statistical analysis of the data, is the experimental unit. The experimental unit is defined as the smallest physical entity to which treatments can be independently assigned (Steel and Torrie 1980) and to which air exchange between test chambers are kept to a minimum. As the number of test chambers/treatments increases, the number of degrees of freedom increases, and, therefore, the width of the confidence interval on a point estimate, such as an LC50, decreases, and the power of a significance test increases (section 12). Because of factors that might affect results within test chambers and results of a test, all test chambers should be treated as similarly as possible. Treatments should be randomly assigned to individual test chamber locations. Assignment of test organisms to test chambers should be nonbiased.

2.2 Types of Tests

2.2.1 A toxicity method is outlined for the euryhaline bivalve, *Mulinia lateralis* (section 11). The manual describes procedures for testing sediments from oligohaline to fully marine environments.

2.3 Test End Points

2.3.1 The end points measured in the toxicity test are survival and growth. Behavior of test organisms should be qualitatively observed daily in all tests (e.g., burrowing).

Section 3

Definitions

3.1 The following terms were defined in Lee (1980), NRC (1989), USEPA (1989b), USEPA-USCOE (1991, 1994), Lee et al. (1992), or ASTM (1993a, 1993b).

3.1.1 Technical Terms

3.1.1.1 *Sediment.* Particulate material that usually lies below water. Formulated particulate material that is intended to lie below water in a test.

3.1.1.2 *Contaminated sediment.* Sediment containing chemical substances at concentrations that pose a known or suspected threat to environmental or human health.

3.1.1.3 *Whole sediment.* Sediment and associated pore water that have had minimal manipulation. The term bulk sediment has been used synonymously with whole sediment.

3.1.1.4 *Control sediment.* A sediment that is essentially free of contaminants and is used routinely to assess the acceptability of a test. Any contaminants in control sediment may originate from the global spread of pollutants and does not reflect any substantial input from local or nonpoint sources. Comparing test sediments to control sediments is a measure of the toxicity of a test sediment beyond inevitable background contamination.

3.1.1.5 *Reference sediment.* A whole sediment near an area of concern used to assess sediment conditions exclusive of material(s) of interest. The reference sediment may be used as an indicator of localized sediment conditions exclusive of the specific pollutant input of concern. Such sediment would be collected near the site of concern and would represent the background conditions resulting from any localized pollutant inputs as well as global pollutant input. This is the manner in which reference sediment is used in dredge material evaluations.

3.1.1.6 *Interstitial water or pore water.* Water occupying space between sediment or soil particles.

3.1.1.7 *Spiked sediment.* A sediment to which a material has been added for experimental purposes.

3.1.1.8 *Reference-toxicity test.* A test conducted in conjunction with sediment tests to determine possible changes in condition of the test organisms. Deviations outside an established normal range indicate a change in the condition of the test organism population. Reference-toxicity tests are most often performed in the absence of sediment.

3.1.1.9 *Clean.* Denotes a sediment or water that does not contain concentrations of test materials that cause apparent stress to the test organisms or reduce their survival.

3.1.1.10 *Overlying water.* The water placed over sediment in a test chamber during a test.

3.1.1.11 *Concentration.* The ratio of weight or volume of test material(s) to the weight or volume of sediment.

3.1.1.12 *No Observable Effect Concentration (NOEC).* The highest concentration of a toxicant to which organisms are exposed in a test that causes no observable adverse effect on the test organisms (i.e., the highest concentration of a toxicant in which the value for the observed response is not significantly different from the controls).

3.1.1.13 *Lowest Observable Effect Concentration (LOEC).* The lowest concentration of a toxicant to which organisms are exposed in a test that causes an adverse effect on the test organisms (i.e., where the value for the observed response is significantly different from the controls).

3.1.1.14 *Lethal concentration (LC).* The toxicant concentration that would cause death in a given percent of the test population. Identical to EC when the observable adverse effect is death. For example, the LC50 is the concentration of toxicant that would cause death in 50% of the test population.

3.1.1.15 *Effect concentration (EC).* The toxicant concentration that would cause an effect in a given percent of the test population. Identical to LC when the observable adverse effect is death. For example, the EC50 is the concentration of toxicant that would cause death in 50% of the test population.

3.1.1.16 *Inhibition concentration (IC).* The toxicant concentration that would cause a given percent reduction in a nonquantal measurement for the test population. For example, the IC25 is the concentration of toxicant that would cause a 25% reduction in growth for the test population, and the IC50 is the concentration of toxicant that would cause a 50% reduction.

3.1.2 Grammatical Terms

3.1.2.1 The words "must," "should," "may," "can," and "might" have very specific meanings in this manual.

3.1.2.2 "Must" is used to express an absolute requirement, that is, to state that a test ought to be designed to satisfy the specified conditions, unless the purpose of the test requires a different design. "Must" is only used in connection with the factors that directly relate to the acceptability of a test.

3.1.2.3 "Should" is used to state that the specified condition is recommended and ought to be met if possible. Although a violation of one "should" is rarely a serious matter, violation of several will often render the results questionable.

3.1.2.4 Terms such as "is desirable," "is often desirable," and "might be desirable" are used in connection with less important factors.

3.1.2.5 "May" is used to mean "is (are) allowed to," "can" is used to mean "is (are) able to," and "might" is used to mean "could possibly." Thus, the classic distinction between "may" and "can" is preserved, and "might" is never used as a synonym for either "may" or "can."

Section 4

Interferences

4.1 General Introduction

4.1.1 Interferences are characteristics of a sediment or sediment test system that can potentially affect test organism survival aside from those related to sediment-associated contaminants. These interferences can potentially confound interpretation of test results in two ways: (1) toxicity is observed in the test when contamination is not present, or there is more toxicity than expected; and (2) no toxicity is observed when contaminants are present at elevated concentrations, or there is less toxicity than expected.

4.1.2 There are three categories of interfering factors: those characteristics of sediments affecting survival independent of chemical concentration (i.e., noncontaminant factors), changes in chemical bioavailability as a function of sediment manipulation or storage, and the presence of indigenous organisms. Although test procedures and test organism selection criteria were developed to minimize these interferences, this section describes the nature of these interferences.

4.1.3 Because of the heterogeneity of natural sediments, extrapolation from laboratory studies to the field can sometimes be difficult (Table 4.1; Burton 1991). Sediment collection, handling, and storage may alter bioavailability and concentration by changing the physical, chemical, or biological characteristics of the sediment. Maintaining the integrity of a field-collected sediment during removal, transport, mixing, storage, and testing is extremely difficult and may complicate the interpretation of effects. Direct comparisons of organisms exposed in the laboratory and in the field would be useful to verify laboratory results. However, spiked sediment may not be representative of contaminated sediment in the field. Mixing time (Stemmer et al. 1990a) and aging (Word et al. 1987, Landrum 1989, Landrum and Faust 1992) of spiked sediment can affect responses of organisms.

4.1.3.1 Laboratory sediment testing with field-collected sediments may be useful in estimating cumulative effects and interactions of multiple contaminants in a sample. Tests with field samples usually cannot discriminate between effects of individual chemicals. Many sediment samples contain a complex matrix of inorganic and organic contaminants with many unidentified compounds. The use of Toxicity Identification Evaluations (TIE) in conjunction with sediment tests with spiked chemicals may provide evidence of causal relationships and can be applied to many chemicals of concern (Ankley and Thomas 1992, Adams et al. 1985). Sediment spiking can also be used to investigate additive, antagonistic, or synergistic effects of specific contaminant mixtures in a sediment sample (Swartz et al. 1988).

4.1.4 Methods that measure sublethal effects are either not available or have not been routinely used to evaluate sediment toxicity (Craig 1984, Dillon and Gibson 1986, Ingersoll and Nelson 1990, Ingersoll 1991, Burton et al. 1994). Most assessments of contaminated sediment rely on acute lethality testing methods (e.g., <10-day; USEPA-USCOE 1977, USEPA-USCOE 1991). Acute lethality tests, such as the 10-day amphipod test, are useful in identifying "hot spots" of sediment contamination, but may not be sensitive enough to evaluate moderately contaminated areas. However, sediment quality assessments using sublethal responses of benthic organisms such as effects on growth and reproduction have been used to successfully evaluate moderately contaminated areas (Scott 1989). Additional methods development is

Table 4.1 ADVANTAGES AND DISADVANTAGES FOR USE OF SEDIMENT TESTS
(Modified from Swartz [1989])

Advantages

- Measure bioavailable fraction of contaminant(s).
- Provide a direct measure of benthic effects, assuming no field adaptation or amelioration of effects.
- Limited special equipment is required.
- Methods are rapid and inexpensive.
- Legal and scientific precedence exists for use; ASTM standard guides are available.
- Measure unique information relative to chemical analyses or benthic community analyses.
- Tests with spiked chemicals provide data on cause-effect relationships.
- Sediment-toxicity tests can be applied to all chemicals of concern.
- Tests applied to field samples reflect cumulative effects of contaminants and contaminant interactions.
- Toxicity tests are amenable to confirmation with natural benthos populations.

Disadvantages

- Sediment collection, handling, and storage may alter bioavailability.
- Spiked sediment may not be representative of field contaminated sediment.
- Natural geochemical characteristics of sediment may affect the response of test organisms.
- Indigenous animals may be present in field-collected sediments.
- Route of exposure may be uncertain and data generated in sediment toxicity tests may be difficult to interpret if factors controlling the bioavailability of contaminants in sediment are unknown.
- Tests applied to field samples may not discriminate effects of individual chemicals.
- Few comparisons have been made of methods or species.
- Only a few chronic methods for measuring sublethal effects have been developed or extensively evaluated.
- Laboratory tests have inherent limitations in predicting ecological effects.

---

needed on chronic sediment testing procedures and culturing of infaunal organisms with a variety of feeding habits including suspension feeders, such as *M. lateralis,* as well as deposit feeders.

4.1.5 Despite the interferences discussed in this section, existing sediment testing methods can be used to provide a rapid and direct measure of effects of contaminants on benthic communities. Laboratory tests with field-collected sediment can also be used to determine temporal, horizontal, or vertical distribution of contaminants in sediment. Most tests can be completed within 2 to 4 weeks. Legal and scientific precedence exist for use of toxicity and bioaccumulation tests in regulatory decision making (e.g., USEPA 1986a). Furthermore, sediment tests with complex contaminant mixtures are important tools for making decisions about the extent of remedial action for contaminated aquatic sites and for evaluating the success of remediation activities.

4.2 Noncontaminant Factors

4.2.1 Results of sediment tests can be used to predict effects that may occur with aquatic organisms in the field as a result of exposure under comparable conditions. Yet, motile organisms might avoid exposure in the field. Photoinduced toxicity caused by ultraviolet (UV) light may be important for some compounds associated with sediment (e.g., polycyclic aromatic hydrocarbons [PAHs]; Davenport and Spacie 1991, Ankley et al. 1994). While natural sunlight contains UV light, lighting typically used to conduct laboratory tests does not include the appropriate spectrum of ultraviolet radiation to photoactivate compounds (Oris and Giesy 1985), and, thus, laboratory tests may not account for toxicity expressed by this mode of action. Lighting can, therefore, affect toxicological responses and is an important experimental variable for photoactivated chemicals.

4.2.2 There are a number of noncontaminant factors that may influence organism survival in these tests. The most important and variable factors include pore water salinity and pore water ammonia. The physico-chemical properties of each test sediment must be within the tolerance limits of the test organism. Tolerance limits of *M. lateralis* for the factors listed above are well defined and are presented in section 11.4 and summarized in Table 11.4. The effects of sediment characteristics can also be extrapolated with regression equations (DeWitt et al. 1988, Ankley et al. 1994) that estimate the proportion of toxicity that may be due to the noncontaminant factor alone.

4.3 Changes in Bioavailability

4.3.1 Sediment toxicity tests are meant to serve as an indicator of contaminant-related toxicity that might be expected under field or natural conditions. Although the tests are not designed to simulate natural conditions, there is concern that contaminant availability in laboratory toxicity test may be different from in-place sediments in the field.

4.3.2 Sediment collection, handling, and storage may alter contaminant bioavailability and concentration by changing the physical, chemical, or biological characteristics of the sediment. These manipulation processes are generally thought to increase availability of organic compounds because of disruption of the equilibrium with organic carbon in the pore water/particle system. Similarly, oxidation of anaerobic sediments increases the availability of certain metals (Di Toro et al. 1990). Because the availability of contaminants may be a function of the degree of manipulation, this manual recommends that handling, storage, and preparation of the sediment for actual testing be as consistent as possible. Although very disruptive of natural sediment physical features, all test sediments should be press-sieved sometime before testing and rehomogenized immediately before introduction to the test chambers if warranted (See section 8.3.1). Press-sieving is performed primarily to remove predatory organisms, large debris, or organisms taxonomically similar to the test species. Certain USEPA program offices may recommend that sediments should not be press-sieved. Also, it may not be necessary to press-sieve sediments if previous experience has demonstrated the absence of potential interferences, including predatory or competitive organisms or large debris, or if large debris or predators can be removed with forceps or other suitable tools. If sediments must be sieved, it may be desirable to perform select analyses (e.g., pore-water metals or DOC, AVS, TOC) on samples before and after sieving to document the influence of sieving on sediment chemistry.

4.3.3 Testing sediments at temperatures different from that in the field might affect contaminant solubility, partitioning coefficients, or other physical and chemical characteristics. Interaction between sediment and overlying water and the ratio of sediment to overlying water may influence bioavailability (Stemmer et al. 1990b).

4.3.4 Depletion of aqueous and sediment-sorbed contaminants resulting from uptake by an organism or test chamber may also influence availability. In most cases, the organism is a minor sink for contaminants relative to the sediment. However, within the burrow of an organism, sediment desorption kinetics may limit uptake rates. Within minutes to hours, a major portion of the total chemical may be inaccessible to the organisms because of depletion of available residues. The desorption of a particular compound from sediment may range from easily reversible (labile; within minutes) to irreversible (nonlabile; within days or months; Karickhoff and Morris 1985). Interparticle diffusion or advection and the quality and quantity of sediment organic carbon can also affect sorption kinetics.

4.3.5 The route of exposure may be uncertain and data from sediment tests may be difficult to interpret if factors controlling the bioavailability of contaminants in sediment are unknown. Bulk-sediment chemical concentrations may be normalized to factors other than dry weight. For example, concentrations of non-ionic organic compounds might be normalized to sediment organic-carbon content (USEPA 1992c) and certain metals normalized to acid volatile sulfides (Di Toro et al. 1990). Even with the appropriate normalizing factors, determination of toxic effects from ingestion of sediment or from dissolved chemicals in the interstitial water can still be difficult (Lamberson and Swartz 1988).

4.3.6 Salinity of the overlying water is an additional factor that can affect the bioavailability of metals. Some metals (e.g., cadmium) are more bioavailable at lower salinities. Therefore, if a sediment sample from a low-salinity location is tested with overlying waters of high salinity, there is the potential that metal toxicity may be reduced.

4.4 Presence of Indigenous Organisms

4.4.1 Indigenous organisms may be present in field-collected sediments. An abundance in the sediment sample of the test organism, or organisms taxonomically similar to the test organism, may make interpretation of treatment effects difficult. The presence of predatory organisms can also adversely affect test organism survival. For example, Redmond and Scott (1989) showed that the polychaete *Nephtys incisa* will consume *Ampelisca abdita* under toxicity test conditions. *Mulinia lateralis* survival in treatment cores in the laboratory was reduced by the presence of two polychaete species. In the field, these polychaetes have been shown to eat juvenile bivalves (Luchenbach 1984). Previous investigators have inhibited the biological activity of sediment with sieving, heat, mercuric chloride, antibiotics, or gamma irradiation (USEPA 1994b, Day et al. 1992). Although further research is needed to determine effects on contaminant bioavailability from treating sediment to remove or destroy indigenous organisms, estuarine and marine sediments must be press-sieved before the start of a sediment toxicity test if the presence of predatory organisms is suspected (see section 8.3.1.1).

Section 5

Health, Safety, and Waste Management

5.1 General Precautions

5.1.1 Development and maintenance of an effective health and safety program in the laboratory requires an ongoing commitment by laboratory management and includes (1) the appointment of a laboratory health and safety officer with the responsibility and authority to develop and maintain a safety program; (2) the preparation of a formal written health and safety plan, which is provided to each laboratory staff member; (3) an ongoing training program on laboratory safety, and (4) regular safety inspections.

5.1.2 This manual addresses procedures that may involve hazardous materials, operations, and equipment, and it does not purport to address all of the safety problems associated with their use. It is the responsibility of the user to establish appropriate safety and health practices, and determine the applicability of regulatory limitations before use. While some safety considerations are included in the manual, it is beyond the scope of the manual to encompass all safety requirements necessary to conduct sediment tests.

5.1.3 Collection and use of sediments may involve substantial risks to personal safety and health. Contaminants in field-collected sediment may include carcinogens, mutagens, and other potentially toxic compounds. Inasmuch as sediment testing is often begun before chemical analyses can be completed, worker contact with sediment needs to be minimized by (1) using gloves, laboratory coats, safety glasses, face shields, and respirators as appropriate; (2) manipulating sediments under a ventilated hood or in an enclosed glove box; and (3) enclosing and ventilating the exposure system. Personnel collecting sediment samples and conducting tests should take all safety precautions necessary for the prevention of bodily injury and illness that might result from ingestion or invasion of infectious agents, inhalation or absorption of corrosive or toxic substances through skin contact, and asphyxiation because of lack of oxygen or presence of noxious gases.

5.1.4 Before sample collection and laboratory work, personnel should determine that all required safety equipment and materials have been obtained and are in good condition.

5.2 Safety Equipment

5.2.1 Personal Safety Gear

5.2.1.1 Personnel should use safety equipment, such as rubber aprons, laboratory coats, respirators, gloves, safety glasses, face shields, hard hats, and safety shoes as appropriate. The degree of protection should vary according to the level of contamination associated with the test sediments. Generally, a higher degree of coverage should be adopted in all aspects of testing sediments that may harbor hazardous levels of compounds. Coverage for testing control or moderately contaminated sediment does not have to be as stringent.

5.2.2 Laboratory Safety Equipment

5.2.2.1 Each laboratory should be provided with safety equipment such as first aid kits, fire extinguishers, fire blankets, emergency showers, and eye fountains.

5.2.2.2 All laboratories should be equipped with a telephone to enable personnel to summon help in case of emergency.

5.3 General Laboratory and Field Operations

5.3.1 Laboratory personnel should be trained in proper practices for handling and using chemicals that are encountered during the procedures described in this manual. Routinely encountered chemicals include acids and organic solvents. Special handling and precautionary guidance in Material Safety Data Sheets should be followed for reagents and other chemicals purchased from supply houses. All containers should be adequately labeled to indicate their contents.

5.3.2 Work with some sediments may require compliance with rules pertaining to the handling of hazardous materials. Personnel collecting samples and performing tests should not work alone.

5.3.3 It is advisable to wash exposed parts of the body with bactericidal soap and water immediately after collecting or manipulating sediment samples.

5.3.4 Strong acids and volatile organic solvents should be used in a fume hood or under an exhaust canopy over the work area.

5.3.5 An acidic solution should not be mixed with a hypochlorite solution because hazardous fumes might be produced.

5.3.6 To prepare dilute acid solutions, concentrated acid should be added to water, not vice versa. Opening a bottle of concentrated acid and adding concentrated acid to water should be performed only in a fume hood.

5.3.7 Use of ground-fault systems and leak detectors is strongly recommended to help prevent electrical shocks. Electrical equipment or extension cords not bearing the approval of Underwriter Laboratories should not be used. Ground-fault interrupters should be installed in all "wet" laboratories where electrical equipment is used.

5.3.8 All containers should be adequately labeled to identify their contents.

5.3.9 Good housekeeping contributes to safety and reliable results.

5.4 Disease Prevention

5.4.1 Personnel handling samples that are known or suspected to contain human wastes should be given the opportunity to be immunized against hepatitis B, tetanus, typhoid fever, and polio.

5.5 Safety Manuals

5.5.1 For further guidance on safe practices when handling sediment samples and conducting toxicity tests, check with the permittee and consult general industrial safety manuals including USEPA (1986b) and Walters and Jameson (1984).

5.6 Pollution Prevention, Waste Management, and Sample Disposal

5.6.1 It is the laboratory's responsibility to comply with the Federal, State and local regulations governing the waste management, particularly hazardous waste identification rules and land disposal restrictions, and to protect the air, water, and land by minimizing and controlling all releases from fume hoods and bench operations. Also, compliance is required with any sewage discharge permits and regulations. For further information on waste management, consult "The Waste Management Manual for Laboratory Personnel," available from the American Chemical Society's Department of Government Relations and Science Policy, 1155 16th Street N.W., Washington, DC 20036.

5.6.2 Guidelines for the handling and disposal of hazardous materials should be strictly followed. The Federal Government has published regulations for the management of hazardous waste and has given the States the option of either adopting those regulations or developing their own. If States develop their own regulations, they are required to be at least as stringent as the Federal regulations. As a handler of hazardous materials, it is your responsibility to know and comply with the pertinent regulations applicable in the State in which you are operating. Refer to the Bureau of National Affairs, Inc. (1986) for the citations of the Federal requirements.

Section 6

Facilities, Equipment, and Supplies

6.1 General

6.1.1 Before a sediment test is conducted in any test facility, it is desirable to conduct a "nontoxicant" test, in which all test chambers contain a control sediment (sometimes called the negative control), and overlying water with no added test material. Survival of the test organism will demonstrate whether facilities, water, control sediment, and handling techniques are adequate to achieve acceptable control survival. Evaluations may also be conducted of the magnitude of the within- and between-chamber variance in a test.

6.2 Facilities

6.2.1 The facility should include separate areas for culturing and testing to reduce the possibility of contamination by test materials and other substances, especially volatile compounds. Culture chambers should not be in a room in which sediment tests are conducted, where stock solutions or where sediments are prepared, or equipment is cleaned. Test chambers may be placed in a temperature controlled recirculating water bath, environmental chamber, or equivalent facility with temperature control. Enclosure of the test systems is desirable to provide ventilation during tests to limit exposure of laboratory personnel to volatile substances.

6.2.2 Light of the quality and illuminance normally obtained in the laboratory is adequate (about 500 to 1000 lux using wide-spectrum fluorescent lights [e.g., cool-white or day-light]) for culturing. Lux is the unit selected for reporting luminance in this manual. Units of foot candles can be converted to units of lux by dividing by 0.093. Units of lux can be converted to units of :E $m_{-2}s_{-1}$ (assuming an average wavelength of 550 nm [:mol_{-2} s_{-1} = W m × 8(nm) × 8.36 × 10^{-3}]) by multiplying by 6.91 × 10^{-3} (ASTM 1994). Luminance should be measured at the surface of the water. *Mulinia lateralis* tests should be conducted under constant light (2000-4000 lux; fluorescent lights suspended directly over the test) so that algae, provided as food, does not drastically reduce the dissolved oxygen in the test chamber through respiration. Illumination should be uniform and continuous during the test period. However, throughout acclimation and holding, a photoperiod from 14L:10D to 16L:8D is recommended.

6.2.3 During rearing, holding, and testing, test organisms should be shielded from external disturbances such as rapidly changing light or pedestrian traffic.

6.2.4 The test facility should be well ventilated and free of fumes. Air used for aeration must be free of oil and fumes. Filters to remove oil, water, and bacteria are desirable. The test facility should be well ventilated and free of fumes. Oil-free air pumps should be used where possible. Particulates can be removed from the air using filters such as BALSTON Grade BX (Balston, Inc., Lexington, MA) or equivalent, and oil and other organic vapors can be removed using activated carbon filters (e.g., BALSTON, C-1 filter) or equivalent. Laboratory ventilation systems should be checked to ensure that return air from chemistry laboratories or sample handling areas is not circulated to culture or testing rooms, or that air from testing rooms does not contaminate culture rooms. Air pressure differentials between rooms should not result in a net flow of potentially contaminated air to sensitive areas through open or loosely fitting doors.

6.3 Equipment and Supplies

6.3.1 Equipment and supplies that contact stock solutions, sediments, or overlying water should not contain substances that can be leached or dissolved in amounts that adversely affect the test organisms. In addition, equipment and supplies that contact sediment or water should be chosen to minimize sorption of test materials from water. Glass, type 316 stainless steel; nylon; and high-density polyethylene, polycarbonate, and fluorocarbon plastics should be used whenever possible to minimize leaching, dissolution, and sorption. High-density plastic containers may be used for holding, acclimation, and culture chambers. These materials should be washed in detergent, acid rinsed, and soaked in flowing water for a week or more before use. Copper, brass, lead, galvanized metal, and natural rubber should not contact overlying water or stock solutions before or during a test. Items made of neoprene rubber and other materials not mentioned above should not be used unless it has been shown that their use will not adversely affect survival, growth, or reproduction of the test organisms.

6.3.2 New lots of plastic products should be tested for toxicity before general use by exposing organisms to them under ordinary test conditions.

6.3.3 General Equipment

6.3.3.1 Environmental chamber or equivalent facility with photoperiod and temperature control (5 to 25 °C).

6.3.3.2 Water purification system capable of producing at least 1 mega-ohm water (USEPA 1993a).

6.3.3.3 Analytical balance capable of accurately weighing to 0.01 mg.

6.3.3.4 Reference weights—Class S, for documenting the performance of the analytical balance(s). The balance(s) should be checked with reference weights that are at the upper and lower ends of the range of the weight values used. A balance should be checked at the beginning of each series of weighings, periodically (such as every tenth weight) during a long series of weighings, and after taking the last weight of a series.

6.3.3.5 Volumetric flasks and graduated cylinders—Class A, borosilicate glass or nontoxic plastic labware, 10 to 4000 mL for making test solutions.

6.3.3.6 Volumetric pipets—Class A, 1 to 100 mL.

6.3.3.7 Serological pipets—1 to 10 mL, graduated.

6.3.3.8 Pipet bulbs and fillers—PROPIPET or equivalent.

6.3.3.9 Droppers, and glass tubing with fire-polished edges, 4 to 6 mm ID—for transferring test organisms.

6.3.3.10 Wash bottles—for rinsing small glassware, instrument electrodes and probes.

6.3.3.11 Glass or electronic thermometers—for measuring water temperature.

6.3.3.12 National Bureau of Standards Certified thermometer (see USEPA Method 170.1; USEPA 1979b).

6.3.3.13 Dissolved oxygen, pH/selective ion, and salinity meters for routine physical and chemical measurements. Unless a test is being conducted to specifically measure the effect of one of these measurements, a portable field-grade instrument is acceptable. A temperature compensated salinity refractometer is useful for measuring salinity of water overlying field collected sediment.

6.3.3.14 Equipment for measuring ammonia (i.e., ammonia-specific probe) is also necessary.

6.3.3.15 See Table 6.1 for a list of additional equipment and supplies.

6.3.4 Test Chambers

6.3.4.1 The test chambers to be used in sediment toxicity tests are 250 mL glass containers (beakers or wide-mouthed jars) with an internal diameter of 10 cm. Each test chamber should be covered to reduce evaporation.

6.3.5 Cleaning

6.3.5.1 All nondisposable sample containers, test chambers, tanks, and other equipment that have come in contact with sediment should be washed after use in the described below to remove surface contaminants.

1. Soak 15 min in tap water and scrub with detergent, or clean in an automatic dishwasher.
2. Rinse twice with tap water.
3. Carefully rinse once with fresh, dilute (10%, V:V) hydrochloric or nitric acid to remove scale, metals, and bases. To prepare a 10% solution of acid, add 10 mL of concentrated acid to 90 mL of deionized water.
4. Rinse twice with deionized water.
5. Rinse once with full-strength, pesticide-grade acetone (use a fume hood or canopy) or muffle at 450 °C for 8 h to remove organic compounds.
6. Rinse three times with deionized water.

6.3.5.2 All test chambers and equipment should be thoroughly rinsed or soaked with the toxicity test diluent water immediately before use in a test.

6.3.5.3 Many organic solvents leave a film that is insoluble in water. A dichromate-sulfuric acid cleaning solution can be used in place of both the organic solvent and the acid (see ASTM 1988), but the solution might attack silicone adhesive and leave chromium residues on glass. An alternative to use of dichromate-sulfuric acid could be to heat glassware for 8 h at 450 °C.

Table 6.1 EQUIPMENT AND SUPPLIES FOR CULTURING AND TESTING *MULINIA LATERALIS*

A. Biological Supplies

Brood stock of test organisms
Algae (e.g., *Isochrysis* sp. T-ISO*, Tetraselmis striata* and *Chaetoceros calcitrans)*

B. Glassware

Culture chambers for brood stock (~20-L tanks)
Culture chambers for larvae and juveniles (2 L glass flat bottomed dishes, beakers)
500-mL bottles
22-L plastic carboys
Test chambers (250-mL glass jar)
Wide-bore pipets (4-to 6-mm ID)
Glass disposable pipets
Graduated cylinders (assorted sizes, 10 mL to 4 L)

C. Instruments and Equipment

Dissecting microscope
Compound microscope
Nylon screens (36, 55, and 500 µm, as well as 1 and 1.5 mm)
Photoperiod timers
Light meter
Temperature controllers
Thermometer
Continuous recording thermometer
Dissolved oxygen meter
pH meter
Ammonia electrode (or ammonia kit)
Salinity meter/temperature compensating salinity refractometer
Drying oven
Desiccator
Hemacytometer
Balance (0.01-mg sensitivity)
Freezer
Refrigerator
Autoclave
Sedgwick-Rafter cell
Downweller (Hadley et al. 1997)

D. Miscellaneous

Ventilation system for test chambers
Air supply and air stones (oil free and regulated)
Glass hole-cutting bits
Glass glue
Aluminum weighing pans (about 2 cm^2)
Fluorescent light bulbs (500-1000 lux)
Deionized water

Table 6.1 EQUIPMENT AND SUPPLIES FOR CULTURING AND TESTING *MULINIA LATERALIS (cont'd.)*

Filtered seawater (0.45 µm)
Air line tubing
White plastic dish pan
Water squirt bottles
Shallow pans (plastic [light-colored], glass, stainless steel)
Dissecting probes

E. Chemicals

Detergent (nonphosphate)
Acetone (reagent grade)
Hexane (reagent grade)
Hydrochloric acid (reagent grade)
F/2 Nutrients
Sodium metasilicate
Formalin
Rose bengal
Cadmium chloride
Copper chloride
Lugol's solution

Section 7

Water, Reagents, and Standards

7.1 Water

7.1.1 Requirements

7.1.1.1 Seawater used to test and culture organisms should be uniform in quality. Acceptable seawater should allow satisfactory survival, growth, or reproduction of the test organisms. Test organisms should not show signs of disease or apparent stress (e.g., discoloration, unusual behavior). If problems are observed in the culturing or testing of organisms, it is desirable to evaluate the characteristics of the water. See USEPA (1993a) and ASTM (1993a) for a recommended list of chemical analyses of the water supply.

7.1.1.2 When deionized water is required, the water-deionizing system should provide sufficient quantity of at least 1 mega-ohm water. If large quantities of high quality deionized water are needed, it may be advisable to supply the laboratory grade water deionizer with preconditioned water from a mixed-bed water treatment system.

7.1.2 Source

7.1.2.1 The source of natural water will depend to some extent on the objective of the test and the test organism that is being used. All natural waters should be obtained from an uncontaminated surface-water source upstream from or beyond the influence of known discharges. Water should be collected at slack high tide, or within 1 h after high tide. Suitable surface water sources should have intakes that are positioned to (1) minimize fluctuations in quality and contamination, (2) maximize the concentration of dissolved oxygen, and (3) ensure low concentrations of sulfide and iron. For estuarine tests, water having a salinity as near as possible to the desired test salinity should be collected from an uncontaminated area. Alternatively, it may be desirable to dilute full strength seawater with an appropriate fresh water source. Sources of fresh water (i.e., 0‰) for dilution include deionized water, uncontaminated well or spring water, or an uncontaminated surface-water source. Municipal water supplies may be variable and may contain unacceptably high concentrations of materials such as copper, lead, zinc, fluoride, chlorine, or chloramines. Chlorinated water should not be used to dilute water utilized for culturing or testing because residual chlorine and chlorine-produced oxidants are toxic to many aquatic organisms. Dechlorinated water should only be used as a last resort for diluting seawater to the desired salinity since dechlorination is often incomplete (ASTM 1993a, USEPA 1993a).

7.1.2.2 For site-specific investigations, it is desirable to have the water-quality characteristics of the overlying water (i.e., salinity) as similar as possible to the site water. For certain applications the experimental design might require use of water from the site where sediment is collected. In estuarine systems, however, the pore water salinity of sediments may not be the same as that of the overlying water at the time of collection (Sanders et al. 1965).

7.1.2.3 Water that might be contaminated with facultative pathogens may be passed through a properly maintained ultraviolet sterilizer equipped with an intensity meter and flow controls or passed through a filter with a pore size of 0.45 µm or less.

7.1.2.4 Natural seawater might need aeration using air stones, surface aerators, or column aerators. Adequate aeration will stabilize pH, bring concentrations of dissolved oxygen and other gases into equilibrium with air, and minimize oxygen demand and concentrations of volatiles. The concentration of dissolved oxygen in source water should be between 90 to 100% saturation to help ensure that dissolved oxygen concentrations are acceptable in test chambers. Natural seawater used for holding/acclimating and testing clams should be filtered (≤1µm) shortly before use to remove suspended particles and organisms. Water used for algal and larval culture should be filtered to 0.45 µm. It may be desirable to aerate dechlorinated water before use as a diluent (e.g., 3 days; J.M. Lazorchak, USEPA, Cincinnati, OH, personal communication).

7.1.2.5 Water that is prepared from natural seawater should be stored in clean, covered containers at 4 °C to eliminate algal growth.

7.1.3 Reconstituted/Synthetic Seawater

7.1.3.1 Although reconstituted water is acceptable, natural seawater is preferable, especially for tests involving chemicals whose bioavailability is affected by seawater chemistry. Reconstituted water is prepared by adding specified amounts of reagent-grade chemicals to high-purity distilled or deionized water (ASTM 1988, USEPA 1991a, 1993a).

7.1.3.2 Suitable salt reagents can be reagent grade chemicals, commercial sea salts, such as Forty Fathoms, Instant Ocean, or HW Marinemix. Preformulated brine (e.g., 60‰ to 100‰), prepared with dry ocean salts or heat-concentrated natural seawater (USEPA 1988) can also be used. A synthetic sea formulation called GP2 is prepared with reagent grade chemicals that can be diluted with a suitable high-quality water to the desired salinity (USEPA 1993c).

7.1.3.3 The suitability and consistency of a particular salt formulation for use in holding and acclimation should be verified by laboratory tests because some formulations can produce unwanted toxic effects or sequester contaminants (Environment Canada 1992).

7.1.3.4 Acceptable high-purity water can be prepared using deionization (at least 1 mega-ohm), distillation, or reverse-osmosis units (section 6.3.3.2; USEPA, 1993a). Test water can also be prepared by diluting natural water with deionized water (Kemble et al. 1994).

7.1.3.5 To obtain the desired holding or acclimation salinity, sea salts or brine can be added to a suitable freshwater or distilled water, or the laboratory's seawater supply may be diluted with a suitable freshwater or distilled water.

7.1.3.7 Salinity, pH, and dissolved oxygen should be measured on each batch of reconstituted water. The reconstituted water should be aerated before use to adjust pH and dissolved oxygen to the acceptable ranges (e.g., section 7.1.3.4.1). Reconstituted seawater should be filtered (≤1µm) shortly before use to remove suspended particles and should be used within 24 h of filtration. USEPA (1993a) recommends using a batch of reconstituted water within a 2-week period.

7.2 Reagents

7.2.1 Test material(s) should be at least reagent grade, unless a test on formulation, commercial product, technical-grade, or use-grade material is specifically needed. Reagent containers should be dated when received from the supplier, and the shelf life of the reagent should not be

exceeded. Working solutions should be dated when prepared and the recommended shelf life should not be exceeded.

7.3 Standards

7.3.1 Appropriate standard methods for chemical and physical analyses should be used when possible. For those measurements for which standards do not exist or are not sensitive enough, methods should be obtained from other reliable sources.

Section 8

Sample Collection, Storage, Manipulation, and Characterization

8.1 Collection

8.1.1 Before the preparation or collection of sediment, a procedure should be established for the handling of sediments that might contain unknown quantities of toxic contaminants (section 5).

8.1.2 A benthic grab (i.e., PONAR, Smith-MacIntyre, Van Veen) or core sampler are preferred sediment samplers because disturbance of sediment samples with these devices is minimized relative to dredge samplers. Although selective subsampling, compositing, and homogenization of sediment samples are necessary for most routine applications addressed by this manual, collection and handling in the field should involve as little disruption as possible. Disruption of sediment samples will cause the loss of sediment integrity and may cause changes in chemical speciation and chemical equilibrium (ASTM 2004). Sediments are spatially and temporally variable (Stemmer et al. 1990a). Replicate samples should be collected to determine variance in sediment characteristics. Sediments should be collected to a depth appropriate for the study objectives. For example, samples collected for evaluations of dredged material should include all sediment to project depth. Surveys of the toxicity of surficial sediment are often based on cores of the upper 2-cm sediment depth.

8.1.3 Exposure to direct sunlight during collection should be minimized, especially if the sediment contains photolytic compounds. Removal of sediment from the sampling device and subsequent allocation to storage containers or homogenization should be accomplished using spoons, trowls, etc. made of, or coated with, inert materials (e.g., Teflon, kynar). Sediment samples should be cooled to 4 °C in the field before returning to the laboratory or shipping (USEPA 1994b). Dry ice can be used to cool samples in the field; however, sediments should never be frozen. Monitors can be used to measure temperature during shipping (e.g., TempTale Temperature Monitoring and Recording System, Sensitech, Inc., Beverly, MA).

8.1.4 For additional information on sediment collection and shipment see ASTM (1990), Burton et al. (1994), and USEPA (1994b).

8.2 Storage

8.2.1 Manipulation or storage can alter bioavailability of contaminants in sediment (Burton and Ingersoll 1994); however, the alterations that occur may not substantially affect toxicity. Although Dillon and Tatem (USCOE, Vicksburg, MS, personal communication) and Ankley and Foe (USEPA, Duluth, MN, unpublished data) found that storage of sediment samples for several months at 4 °C did not result in significant changes in chemistry or toxicity, others have demonstrated that changes in spiked sediment can occur within days to weeks (e.g., Burton 1991; Stemmer et al. 1990a). Sediments primarily contaminated with non-ionic, nonvolatile organic compounds will probably change little during storage because of their relative resistance to biodegradation and sorption to solids. However, metals and metalloids may be affected by changing redox, oxidation, or microbial metabolism (such as with arsenic, selenium, mercury, lead, and tin; all of which are methylated by a number of bacteria and fungi). Metal contaminated sediments may need to be tested relatively soon after collection with as little manipulation as possible (Burton and Ingersoll 1994).

8.2.2 Given that the contaminants of concern and the influencing sediment characteristics are not always known a priori, it is desirable to hold sediments in the dark at 4 °C and start testing soon after collection from the field. Recommended sediment holding time ranges from less than 2 (ASTM 1993a) to less than 8 weeks (USEPA-USCOE 1994). If whole-sediment tests are started after 2 weeks of collection, it may be desirable to conduct additional characterizations of sediment to evaluate possible effects of storage on sediment. For example, concentrations of contaminants of concern could be measured in pore water within 2 weeks from sediment collection and at the start of the sediment test (Kemble et al. 1994). Ingersoll et al. (1993) recommend conducting a toxicity test with pore water within 2 weeks of sediment collection and at the start of the sediment test. Freezing and longer storage might further change sediment properties such as grain size or partitioning and should be avoided (ASTM 1990; Schuytema et al. 1989; K.E. Day, Environment Canada, Burlington, Ontario, personal communication). Sediment should be stored with no air over the sealed samples (no head space) at 4 °C before the start of a test (Shuba et al. 1978, ASTM 1990). Sediment may be stored in containers constructed of suitable materials as outlined in section 6. Care should be taken to avoid contaminating the samples during storing, handling, and sampling.

8.3 Manipulation

8.3.1 Homogenization

8.3.1.1 Sediment samples tend to settle during shipment. Water above the sediment should not be discarded, but should be mixed back into the sediment during homogenization. If warranted, sediment samples should be press-sieved through a 1 or 2 mm mesh stainless steel screen to remove indigenous and/or predatory organisms, large debris, or organisms taxonomically similar to the test species. Certain USEPA program offices may recommend that sediments should not be press-sieved. Also, it may not be necessary to press-sieve sediments if previous experience has demonstrated the absence of potential interferences, including predatory or competitive organisms or large debris, or if large debris or predators can be removed with forceps or other suitable tools. If sediments must be sieved, it may be desirable to perform select analyses (e.g., pore-water metals or DOC, AVS, TOC) on samples before and after sieving to document the influence of sieving on sediment chemistry.

8.3.1.2 If sediment is collected from multiple field samples, the sediment can be pooled and mixed using stirring or a rolling mill, feed mixer, or other suitable apparatus (see ASTM 1990). It is preferable to homogenize sediments by gentle hand mixing. Although potentially disruptive, large numbers of sediments may demand the use of a mechanical aid. Mechanical homogenization of sediment can be accomplished using a modified 30-cm bench-top drill press (Dayton Model 3Z993) or a variable-speed hand-held drill outfitted with a stainless steel auger (diameter 7.6 cm, overall length 38 cm, auger bit length 25.4 cm; Augers Unlimited, Exton, PA; Kemble et al. 1994). These procedures could also be used to mix test sediment with a control sediment in dilution experiments.

8.3.2 Sediment Spiking

8.3.2.1 Test sediment can be prepared by manipulating the properties of a control sediment. Mixing time (Stemmer et al. 1990a) and aging (Word et al. 1987, Landrum 1989; Landrum and Faust 1992) of spiked sediment can affect responses. Many studies with spiked sediment are often started only a few days after the chemical has been added to the sediment. This short time period may not be long enough for sediments to equilibrate with the spiked chemicals. Consistent spiking procedures should be followed in order to make interlaboratory comparisons. It

is recommended that spiked sediment be aged at least 1 month before starting a test; however equilibration for some chemicals may not be achieved for long periods of time.

8.3.2.1.1 The cause of sediment toxicity and the magnitude of interactive effects of contaminants can be estimated by spiking a sediment with chemicals or complex waste mixtures (Lamberson et al. 1992). Sediments spiked with a range of concentrations can be used to generate either point estimates (e.g., LC50) or a minimum concentration at which effects are observed (lowest observable effect concentration; LOEC). The influence of sediment physico-chemical characteristics on chemical toxicity can also be determined with sediment-spiking studies (Adams et al. 1985).

8.3.2.2 The test material(s) should be at least reagent grade, unless a test on formulation commercial product, technical-grade, or use-grade material is specifically needed. Before a test is started, the following should be known about the test material: (1) the identity and concentration of major ingredients and impurities; (2) water solubility in test water; (3) estimated toxicity to the test organism and to humans; (4) if the test concentration(s) are to be measured, the precision and bias of the analytical method at the planned concentration(s) of the test material; and (5) recommended handling and disposal procedures.

8.3.2.2.1 Organic compounds are generally added in the dry form or coated on the inside walls of the mixing container (Ditsworth et al. 1990). Metals are generally added in an aqueous solution (USEPA 1994b, Burton et al. 1994b, Carlson et al. 1991, Di Toro et al. 1990). Concentrations of the chemical in the pore water and in whole sediment should be monitored at the beginning and end of a test.

8.3.2.3 Use of a solvent other than water should be avoided if possible. Addition of organic solvents may dramatically influence the concentration of dissolved organic carbon in pore water (G.T. Ankley, USEPA, Duluth, MN, personal communication). If an organic solvent must be used, both a solvent-control and a negative-control sediment must be included in a test. The solvent in the sediment should be at a concentration that does not affect the test organism. The solvent control must contain the highest concentration of solvent present and must be from the same batch used to make the stock solution (see ASTM 1988). The same concentration of solvent should be used in all treatments. If an organic solvent is used as a carrier, it is possible to perform successive washes of sediment to remove most of the solvent while leaving the compound of study (Harkey et al. 1994).

8.3.2.4 If the concentration of solvent is not the same in all test solutions that contain test material, a solvent test should be conducted to determine whether survival and/or growth of the test organisms is related to the concentration of the solvent.

8.3.2.4.1 If the test contains both a negative control and a solvent control, the survival and/or growth of the organisms tested should be compared. If a statistically significant difference is detected between the two controls, only the solvent control may be used for meeting the acceptability of the test and as the basis for calculating results. The negative control might provide additional information on the general health of the organisms tested. If no statistically significant difference is detected, the data from both controls should be used for meeting the acceptability of the test and as the basis for calculating the results (ASTM 1992).

8.3.2.5 Test Concentration(s) for Laboratory Spiked Sediments

8.3.2.5.1 If a test is intended to generate an LC50, the selected test concentrations should bracket the predicted LC50. The prediction might be based on the results of a test on the same or a similar test material with the same or a similar test organism. The LC50 of a particular compound may vary depending on physical and chemical sediment characteristics. If a useful prediction is not available, it is desirable to conduct a range-finding test in which the organisms are exposed to a control and three or more concentrations of the test material that differ by a factor of 10. Results from water-only tests could be used to establish concentrations to be tested in a whole-sediment test based on predicted pore-water concentrations (Di Toro et al. 1991).

8.3.2.5.2 Bulk-sediment chemical concentrations might be normalized to factors other than dry weight. For example, concentrations of nonpolar organic compounds might be normalized to sediment organic-carbon content and simultaneously extracted metals might be normalized to acid volatile sulfides (Di Toro et al. 1990, 1991).

8.3.2.5.3 In some situations it might be necessary to only determine whether a specific concentration of test material is toxic to the test organism, or whether adverse effects occur above or below a specific concentration. When there is interest in a particular concentration, it might only be necessary to test that concentration and not to determine an LC50.

8.3.2.6 Addition of test material(s) to sediment may be accomplished using various methods, such as (1) a rolling mill (preferred), (2) a feed mixer, or (3) hand mixing (ASTM 1990). Modifications of the mixing techniques might be necessary to allow time for a test material to equilibrate with the sediment. Mixing time of spiked sediment should be limited from minutes to a few hours and temperature should be kept low to minimize potential changes in the physico-chemical and microbial characteristics of the sediment (ASTM 1990, Burton et al. 1994, USEPA 1994b). Duration of contact between the chemical and sediment can affect partitioning and bioavailability (Word et al. 1987). Care should be taken to ensure that the chemical is thoroughly and evenly distributed in the sediment. Analyses of sediment subsamples is advisable to determine the degree of mixing homogeneity (Ditsworth et al. 1990). Moreover, results from sediment-spiking studies should be compared with the response of test organisms to chemical concentrations in natural sediments (Lamberson and Swartz 1988).

8.4 Characterization

8.4.1 All sediments should be characterized and at least the following determined: salinity, pH, and ammonia of the pore water; organic carbon content (total organic carbon, TOC); particle size distribution (percent sand, silt, clay); and percent water content (ASTM 1993a, Plumb 1981). Salinity of sediment pore water should be measured on the supernatant of an aliquot of the sediment using a refractometer or conductivity meter. See section 8.4.4.7 for methods to isolate pore water.

8.4.2 Other analyses on sediments might include biological oxygen demand, chemical oxygen demand, cation exchange capacity, Eh, total inorganic carbon, total volatile solids, acid volatile sulfides, metals, synthetic organic compounds, oil and grease, petroleum hydrocarbons, and interstitial water analyses.

8.4.3 Macrobenthos may be quantified by subsampling the field-collected sediment. If direct comparisons are to be made, subsamples for toxicity testing should be collected from the same

sample for analysis of sediment physical and chemical characterizations. Qualitative descriptions of the sediment may include color, texture, presence of hydrogen sulfide, and presence of indigenous organisms. Monitoring the odor of sediment samples should be avoided because of potential hazardous volatile contaminants. It may be desirable to describe color and texture gradients that occur with sediment depth.

8.4.4 Analytical Methods

8.4.4.1 Chemical and physical data should be obtained using appropriate standard methods whenever possible. For those measurements for which standard methods do not exist or are not sensitive enough, methods should be obtained from other reliable sources.

8.4.4.2 The precision, accuracy, and bias of each analytical method used should be determined in the appropriate matrix: sediment, water, and tissue. Reagent blanks and analytical standards should be analyzed and recoveries should be calculated.

8.4.4.3 Concentration of spiked test material(s) in sediment, interstitial water, and overlying water should be measured as often as practical during a test. If possible, the concentration of the test material in overlying water, interstitial water and sediments should be measured at the start and end of a test. Measurement of test material(s) degradation products might also be desirable.

8.4.4.4 Separate chambers should be set up at the start of a test and destructively sampled during and at the end of the test to monitor sediment chemistry. Test organisms might be added to these extra chambers depending on the objective of the study.

8.4.4.5 Measurement of test material(s) concentration in water can be accomplished by pipeting water samples from about 1 to 2 cm above the sediment surface in the test chamber. Overlying water samples should not contain any surface debris, any material from the sides of the test chamber, or any sediment.

8.4.4.6 Measurement of test material(s) concentration in sediment at the end of a test can be taken by siphoning most of the overlying water without disturbing the surface of the sediment, then removing appropriate aliquots of the sediment for chemical analysis.

8.4.4.7 A variety of procedures have been used to isolate interstitial water including centrifugation, filtration, pressure, or by using an interstitial water sampler; however, centrifugation without filtration is the recommended procedure. Filtration may reduce concentrations of materials in interstitial water (Schults et al. 1992). We recommend centrifuging the sediment in Teflon centrifuge tubes for 1 h at 12,000*g to 15,000*g at 4 °C. Care should be taken to ensure that contaminants do not transform, degrade, or volatilize during isolation or storage of the interstitial water sample.

Section 9

Quality Assurance and Quality Control

9.1 Introduction

9.1.1 Developing and maintaining a laboratory quality assurance (QA) program requires an ongoing commitment by laboratory management and also includes the following: (1) appointment of a laboratory quality assurance officer with the responsibility and authority to develop and maintain a QA program; (2) preparation of a Quality Assurance Project Plan with Data Quality Objectives; (3) preparation of written descriptions of laboratory Standard Operating Procedures and (SOPs) for test organism culturing, testing, instrument calibration, sample chain-of-custody, laboratory sample tracking system; and (4) provision of adequate qualified technical staff and suitable space and equipment to assure reliable data. Additional guidance for QA can be obtained in USEPA (1989b).

9.1.2 QA practices within a testing laboratory should address all activities that affect the quality of the final data, such as (1) sediment sampling and handling, (2) the source and condition of the test organisms, (3) condition and operation of equipment, (4) test conditions, (5) instrument calibration, (6) replication, (7) use of reference toxicants, (8) record keeping, and (9) data evaluation.

9.1.3 Quality Control (QC) practices, on the other hand, consist of the more focused, routine, day-to-day activities carried out within the scope of the overall QA program. For more detailed discussion of quality assurance, and general guidance on good laboratory practices related to testing see FDA (1978), USEPA (1979a; 1980a,b; 1990c,d; 1993a,b,c) DeWoskin (1984), and Taylor (1987).

9.2 Performance-Based Criteria

9.2.1 USEPA Environmental Monitoring Management Council (EMMC) recommended the use of performance-based methods in developing standards for chemical analytical methods (Williams 1993). Performance-based methods were defined by EMMC as a monitoring approach that permits the use of appropriate methods that meet preestablished demonstrated performance standards. Minimum required elements of performance, such as precision, reproducibility, bias, sensitivity, and detection limits should be specified and the method should be demonstrated to meet the performance standards.

9.2.2 Therefore, a performance-based criteria approach was selected as the preferred method through which individual laboratories should evaluate culture methods or the quality of field-collected organisms rather than by control-based criteria. This method was chosen to allow each laboratory to optimize culture methods, determine the quality of field-collected organisms, and minimize effects of test organism health on the reliability and comparability of test results. See Table 11.3 for a listing of culture performance criteria, quality performance criteria for field collected organisms, as well as criteria for result evaluation.

9.3 Facilities, Equipment, and Test Chambers

9.3.1 Separate test organism culturing and testing areas should be provided to avoid loss of cultures because of cross-contamination. Ventilation systems should be designed and operated to prevent recirculation or leakage of air from chemical analysis laboratories or sample storage and preparation areas into test organism culturing or sediment testing areas and from sediment testing laboratories and sample preparation areas into culture rooms.

9.3.2 Equipment for temperature control should be adequate to maintain recommended test-water temperatures. Recommended materials should be used in the fabrication of the test equipment that comes in contact with the sediment or overlying water.

9.3.3 Before a sediment test is conducted in a new facility, a "noncontaminant" test should be conducted in which all test chambers contain a control sediment and overlying water. This information is used to demonstrate that the facility, control sediment, water, and handling procedures provide acceptable responses of test organisms (section 9.14).

9.4 Test Organisms

9.4.1 The organisms should appear healthy, behave normally, feed well, and have low mortality in cultures, during holding, and in test controls. Acceptable mortality rate of larvae through metamorphosis to juveniles is 50%, and for juveniles and adult clams are 30% and 20%, respectively. Test organisms should be positively identified to species. Test organisms should not show signs of disease or apparent stress (e.g., discoloration, unusual behavior).

9.5 Water

9.5.1 The quality of water used for organism culturing and testing is extremely important. Overlying water used in culturing, holding, acclimation, and testing organisms should be uniform in quality. Acceptable water should allow satisfactory survival and growth of the test organisms. See section 7 for guidance on selection and preparation of high-quality test water.

9.6 Sample Collection and Storage

9.6.1 Sample holding times and temperatures should conform to conditions described in section 8.

9.7 Test Conditions

9.7.1 It is desirable to measure temperature continuously in at least one chamber during the each test. Temperatures should be maintained within the limits specified for each test. Dissolved oxygen, salinity, ammonia, and pH should be checked as prescribed in section 11.3.

9.8 Quality of Test Organisms

9.8.1 If test organisms are obtained from culture, monthly reference-toxicity tests should be conducted on all test organisms using procedures outlined in section 9.16. If reference-toxicity tests are not conducted monthly, the lot of organisms used to start a sediment test must be evaluated using a reference toxicant.

9.8.2 The testing laboratory will have to conduct five initial reference toxicity tests with *Mulinia lateralis* (section 9.14.1). Subsequently, the quality of test organisms obtained from an outside source, regardless of whether they are from culture or collected from the field, must be verified by conducting a reference-toxicity test concurrently with the sediment test and comparing the results with the five initial tests conducted. For cultured organisms, the supplier should provide data with the shipment describing the history of the sensitivity of organisms from the same source culture. For field-collected organisms, the supplier should provide data with the shipment describing the collection location, the time and date of collection, the water salinity and temperature at the time of collection, and collection site sediment for holding and acclimation purposes.

9.8.3 The supplier should also certify the species identification of the test organisms and provide the taxonomic references or name(s) of the taxonomic expert(s) consulted.

9.9 Quality of Food

9.9.1 Problems with the nutritional suitability of the food will be reflected in the survival, growth, or reproduction of *Mulinia lateralis.*

9.10 Test Acceptability

9.10.1 For the test results to be acceptable, survival at 10 days must equal or exceed 80% in the control sediment and growth must have increased at least by a factor of two. See Table 11.3 for additional requirements for acceptability of the tests.

9.10.2 An individual test may be conditionally acceptable if temperature, dissolved oxygen, and other specified conditions fall outside specifications, depending on the degree of the departure and the objectives of the tests (see Table 11.3). The acceptability of a test will depend on the experience and professional judgment of the laboratory analyst and the reviewing staff of the regulatory authority. Any deviation from test specifications should be noted when reporting data from a test.

9.11 Analytical Methods

9.11.1 All routine chemical and physical analyses for culture and testing water, food, and sediment should include established quality assurance practices outlined in USEPA methods manuals (USEPA 1979a,b; 1993a).

9.11.2 Reagent containers should be dated when received from the supplier, and the shelf life of the reagent should not be exceeded. Working solutions should be dated when prepared, and the recommended shelf life should not be exceeded.

9.12 Calibration and Standardization

9.12.1 Instruments used for routine measurements of chemical and physical characteristics such as pH, dissolved oxygen, temperature, and salinity should be calibrated before use each day according to the instrument manufacturer's procedures as indicated in the general section on quality assurance (see USEPA Methods 150.1, 360.1, 170.1, and 120.1, USEPA 1979b). Calibration data should be recorded in a permanent log.

9.12.2 A known-quality water should be used in the analyses of each batch of water samples (e.g., water salinity, hardness, alkalinity).

9.13 Replication and Test Sensitivity

9.13.1 The sensitivity of sediment tests will depend in part on the number of replicates per treatment, the significance level selected, and the type of statistical analysis. If the variability remains constant, the sensitivity of a test will increase as the number of replicates is increased. The minimum recommended number of replicates varies with the objectives of the test and the statistical method used for analysis of the data (section 12).

9.14 Demonstrating Acceptable Performance

9.14.1 It is the responsibility of a laboratory to demonstrate its ability to obtain consistent, precise results with reference toxicants before it performs sediment tests (see Section 9.16). Intralaboratory precision, expressed as a coefficient of variation, of the range for each type of test to be used in a laboratory should be determined by performing five or more tests with different batches of test organisms, using the same reference toxicant, at the same concentrations, with the same test conditions (e.g., the same test duration, type of water, age of test organisms) and same data analysis methods. A reference toxicant concentration series (0.5 or higher) should be selected that will consistently provide partial mortalities at two or more concentrations of the test chemical (section 12).

9.14.2 Before conducting tests with contaminated sediment, the laboratory should demonstrate its ability to conduct tests by conducting five exposures in control sediment as outlined in Table 11.1. It is recommended that these five exposures with control sediment be conducted concurrently with the five reference toxicity tests described in section 9.14.1.

9.14.3 Laboratories should demonstrate that their personnel are able to recover an average of at least 80% of the organisms from whole sediment. For example, test organisms could be added to control sediment or test sediments and recovery could be determined after 1 h (Tomasovic et al. 1994).

9.15 Documenting Ongoing Laboratory Performance

9.15.1 Satisfactory laboratory performance on a continuing basis is demonstrated by conducting monthly reference-toxicity tests with the test organism. For a given test organism, successive tests should be performed with the same reference toxicant, at the same concentrations, in the same type of water, generating LC50s using the same data analysis method (section 13).

9.15.2 Outliers, which are data falling outside the control limits, and trends of increasing or decreasing sensitivity are readily identified. If the reference toxicity datum from a given test falls outside the "expected" range (e.g., ±2 SD), the sensitivity of the organisms and the credibility of the test results are suspect. In this case, the test procedure should be examined for defects and should be repeated with a different batch of test organisms.

9.15.3 A sediment test may be acceptable if specified conditions of a reference toxicity test fall outside the expected ranges (section 9.10.2). Specifically, a sediment test should not automatically be judged unacceptable if the LC50 for a given reference toxicity test falls outside the expected range or if mortality in the control of the reference toxicity test exceeds 20%. All

the performance criteria outlined in Table 11.3 must be considered when determining the acceptability of a sediment test. The acceptability of the sediment test would depend on the experience and judgment of the investigator and the regulatory authority.

9.15.4 Performance should improve with experience, and the control limits should gradually narrow, as the statistics stabilize. However, control limits of ±2 SD, by definition, will be exceeded 5% of the time, regardless of how well a laboratory performs. For this reason, good laboratories that develop very narrow control limits may be penalized if a test result which falls just outside the control limits is rejected de facto. The width of the control limits should be considered in decisions regarding rejection of data (section 13).

9.16 Reference Toxicants

9.16.1 Reference-toxicity tests should be conducted in conjunction with sediment tests to determine possible changes in condition of a test organism (Lee 1980). Water-only reference-toxicity tests on cultured organisms should be conducted monthly and should be performed on each batch of field-collected organisms used for testing. Deviations outside an established normal range may indicate a change in the condition of the test organism population. Results of reference-toxicity tests also enable interlaboratory comparisons of test organism sensitivity.

9.16.2 Reference toxicants such as cadmium (available as cadmium chloride, $CdCl_2$), and copper, available as copper chloride ($CuCl_2$), are suitable for use. No one reference toxicant can be used to measure the condition of test organisms in respect to another toxicant with a different mode of action (Lee 1980). However, it may be unrealistic to test more than one or two reference toxicants routinely.

9.16.3 Test conditions for conducting reference-toxicity tests with *Mulinia lateralis* are outlined in Table 9.1.

9.17 Record Keeping

9.17.1 Proper record keeping is important. A complete file should be maintained for each individual sediment test or group of tests on closely related samples. This file should contain a record of the sample chain-of-custody; a copy of the sample log sheet; the original bench sheets for the test organism responses during the sediment test(s); chemical analysis data on the sample(s); control data sheets for reference toxicants; detailed records of the test organisms used in the test(s), such as species, source, age, date of receipt, and other pertinent information relating to their history and health; information on the calibration of equipment and instruments; test conditions used; and results of reference toxicant tests. Laboratory data should be recorded immediately to prevent the loss of information or inadvertent introduction of errors into the record. Original data sheets should be signed and dated by the laboratory personnel performing the tests. For additional detail see section 12.

Table 9.1 RECOMMENDED TEST CONDITIONS FOR CONDUCTING REFERENCE-TOXICITY TESTS WITH *MULINIA LATERALIS*

Parameter	Conditions
1. Test type:	Water-only or spiked sand (50 mL with 150 mL overlying seawater)
2. Dilution series:	Control and at least 5 test concentrations (0.5 dilution factor)
3. Toxicant:	Cd, Cu
4. Temperature:	22 ±2 °C
5. Salinity:	Although 30 ±2‰ is the standard, the test salinity can be adapted to lower salinities as long as these fall within the tolerance range of the species (see section 4.3.6 for effects on metal bioavailability).
6. Light Quality:	Chambers should be kept under fluorescent light. (2000-4000 lux)
7. Photoperiod:	24-h light
8. Renewal of water:	None
9. Size of test organisms:	1 to 1.5 mm
10. Test chamber:	250-mL glass jar
11. Volume of water	150 mL (minimum)
12. Number of replicate chambers/treatment:	3 (minimum)
13. Number of organisms/chamber:	10 (minimum)
14. Aeration:	None
15. Dilution water:	Culture water, surface water, site water, or reconstituted water
16. Water quality:	Salinity, pH, and dissolved oxygen at beginning and end of test; temperature daily
17. Test duration:	96 h
18. End point:	Survival (LC50) and growth (EC50)
19. Test acceptability:	80% control survival

Section 10

Collection, Culture, and Maintenance of Test Organisms

10.1 Life History

10.1.1 *Mulinia lateralis,* the coot clam or baby surf clam, is a mollusc of the class Bivalvia. It is found in shallow (<32-m), soft, sand/clay sediments of the Atlantic and Gulf of Mexico coasts from Canada to Mexico (Rhodes et al. 1972). *Mulinia lateralis* is euryhaline and is generally found in salinities between 7 and 34‰, although it has been reported in waters that range from 1.4 to 75‰ (Calabrese 1969). It has an average life span of 2 years and, in the field, is characterized by high rates of growth and reproduction followed by episodic mass mortality (Shumway and Newell 1984). Such mortality may be due to predation (Virnstein 1977, Holland et al. 1980, Luchenbach 1984) or metabolic stress when energetic requirements exceed food availability (Kennedy and Mihurski 1972, Shumway and Newell 1984). It is the food of many bottom dwelling and bottom feeding animals such as fish, starfish, oyster drills, and water fowl. This filter-feeding bivalve feeds primarily on phytoplankton, although there is some evidence of suspension feeding (Frankenberg et al. 1967). In Long Island Sound, most gametogenic activity does not begin until the water temperature reaches 7 °C, then increases further as the water temperature increases. Gametogenesis continues into June and July. Spawning begins in early July and continues through September. More than one generation is produced in a single year (Calabrese 1970). In more southern habitats, *M. lateralis* has a much more extended reproductive season (Hanks 1968, Williams et al. 1986). In the laboratory, this bivalve has been raised from fertilized egg to reproducing adult, when fed naturally available algae, within 60 days (Calabrese 1970, 1974).

10.2 Obtaining Adults for Broodstock

10.2.1 Field Collections

10.2.1.1 Field collection is presently the most common method for obtaining brood stock. The availability of ripe adults will vary seasonally depending on latitude. The collection site chosen should be one for which the presence of abundant organisms has been demonstrated previously, and identification of the species has been confirmed taxonomically (e.g., Smith 1964).

Collection areas should have no, or minimal contamination. Because different populations may exhibit different sensitivities to contaminants, broodstock used to obtain juveniles for testing must be obtained from the same locality.

10.2.2 Collection Methods

10.2.2.1 Although it is desirable to collect ripe adults from the field, juveniles can be collected, and held until they are in reproductive state in the laboratory. Sediment samples from sites inhabited by *M. lateralis* can be collected with a small dredge or grab (e.g., PONAR, Smith-McIntyre, or Van Veen). *M. lateralis* can be isolated from collection site sediment by gentle sieving onto a 2-mm stainless steel sieve.

10.2.2.2 Collection-site water should be used to sieve sediment in the field. A 2-cm thick layer of sieved collection site sediment should be placed in transport containers, and this sediment covered with collection-site water. Detritus and predators recovered by sieving should be removed, and the collected clams should be gently washed into the transport containers with collection-site water.

10.2.2.3 The salinity and temperature of surface and bottom seawater at the collection site should be measured and recorded. An adequate portion of collection-site sediment should be returned with the clams to serve as both laboratory holding sediment and for use as control sediment in the toxicity test.

10.2.2.4 All apparatus used for collecting, sieving, and transporting clams and control-site sediment should be clean and made of nontoxic material. They should be marked, "Live only," and must never be used for working with formalin or any other toxic materials and should be stored separately from the aforementioned. The containers and other collection apparatus should be cleaned and rinsed with deionized water, dechlorinated laboratory water, reconstituted seawater, or natural seawater from the collection site or an uncontaminated source before use.

10.2.2.5 During transport to the laboratory, clams should be kept in sieved collection-site sediment at or below the collection-site temperature. Containers of clams and sediment should be transported to the laboratory in coolers with ice packs, and the water in the containers of clams should be aerated if transport time exceeds 1 h.

10.2.3 Purchasing *M. lateralis*

10.2.3.1 *M. lateralis* can be purchased from suppliers or may be obtained from laboratories with existing cultures. In either case shipping may be required. Adults are available from Marine Biological Laboratory, Department of Marine Resources, Woods Hole, MA 02543, phone (508)289-7375, e-mail specimens@mbl.edu. Although juveniles are not currently available from suppliers, they may be in the future.

10.2.3.2 Shipping Methods

10.2.3.2.1 Adult and juvenile individuals of this species have been routinely shipped from the collection site to the laboratory for use as broodstock or for testing. It is critical that standard, demonstrated shipping methods are utilized to ensure the health of the organisms and consistency of test results.

10.2.3.2.2 Adult, field-collected clams should be shipped within 24 h of collection. Approximately 20 individuals can be placed in 500-mL containers containing a mixture of *Isochrysis* sp. (T-ISO) and *T. striata.* The algae should be aerated prior to addition to the shipping containers, which should then be immediately sealed and placed in a cooler for shipment. Juvenile clams should be shipped in a similar manner although they can also be shipped in a small amount of sediment. Small plastic "sandwich" containers (approximately 500 mL) with tightly fitting lids may be used. A 1-cm layer of sieved collection site sediment (fine silt or sand) is added to each container and then filled three-quarters full with collection-site seawater.

10.2.3.3 Performance Criteria for Shipped Clams

10.2.3.3.1 The process of ensuring the availability of healthy clams begins when the animals arrive in the laboratory from the supplier. The shipping containers should arrive intact, and the

temperature of water or sediment in shipping containers should be between 4 and 10 °C. Mortality among the shipped animals should not exceed 5% and active burrowing should occur within hours of placement over sediment. Information on physical parameters of the collection site, including at least temperature and salinity, should be provided by the supplier. Finally, a quantity of collection site sediment should be included as substratum for clams during the acclimation period, and for use as control sediment in toxicity testing.

10.3 Holding and Acclimation of Adults (Broodstock)

10.3.1 Density. *M. lateralis* broodstock may be held and acclimated (if necessary) in 20-L aquaria that contain approximately 2.5 cm of collection site sediment that has been sieved through a 0.5-mm mesh screen, or sand. Approximately 75 adult clams should be added to each 20-L aquarium. Juvenile clams can be raised at much higher densities (>100/tank). Aeration, provided through a pipet, should be gentle and constant.

10.3.2 Duration. *M. lateralis* broodstock and juveniles can be held in the laboratory for several months to more than a year. Temperature and salinity should be measured daily. Dissolved oxygen and pH should be measured in the holding containers at least at the start of the acclimation period and weekly thereafter.

10.3.3 Temperature. *M. lateralis* can tolerate temperatures between 0 °C and 29 °C. The holding temperature for the broodstock should be maintained at 23 °C, as temperatures below 20 °C may not foster adequate reproductive rates. If the temperature (at the time of collection or shipping) is very different from the holding temperature (23 °C), the animals can be acclimated at a rate of 3 °C per day.

10.3.4 Salinity. The recommended holding salinity for *M. lateralis* may range between 27‰ and 32‰, with daily changes of no more than 3‰, except in cases where tests will be conducted with lower salinity. If the salinity at the collection site is significantly different from this, acclimation should be at a rate of no greater than 3‰ per day.

10.3.5 Lighting. Lighting should provide a photoperiod of 16L:8D or 14L:10D throughout the holding and acclimation period using fluorescent lights at a light intensity of 500 to 1000 lux.

10.3.6 Water

10.3.6.1 Provided that it is acceptable for holding adults, either an uncontaminated supply of natural seawater or reconstituted seawater can be used for holding and acclimation (section 7). At a minimum, healthy clams must exhibit acceptable survival in holding water and must not exhibit signs of stress, such as unusual behavior or changes in appearance.

10.3.6.2 If natural seawater is used, it should be obtained from an uncontaminated area known to support a healthy, reproducing population of the test species or comparably sensitive species. Reconstituted seawater is prepared by adding commercially available sea salts to water from a suitable source, in quantities sufficient to provide the desired salinity. Preformulated brine (e.g., 60% to 90%)—prepared with dry sea salts or heat-concentrated natural seawater—can also be used. The suitability and consistency of a particular salt formulation for use in holding and acclimation should be verified by laboratory tests because some formulations can produce unwanted toxic effects or sequester contaminants (Environment Canada 1992). Suitable sources of water used for preparing reconstituted seawater include deionized or distilled water or an uncontaminated natural surface or groundwater. Chlorinated water must

never be used because residual chlorine and chlorine-produced oxidants are highly toxic to many aquatic animals. Because dechlorination is often incomplete, dechlorinated municipal drinking water should be used only as a last resort. Reconstituted seawater should be intensively aerated for 2 weeks before use (Environment Canada 1992).

10.3.6.3 Assessments of the quality of the water used for holding and for preparing reconstituted seawater should be performed as frequently as required to document acceptability. Analyses of variables including salinity, temperature, pH, and dissolved oxygen, are recommended. Seawater used for holding and acclimating clams should be filtered ($\leq$1 µm) shortly before use to remove suspended particles and organisms. Holding/acclimation water prepared from natural seawater should be used within 2 days of filtration/sterilization, whereas reconstituted seawater should be used within 24 h of filtration/sterilization.

10.3.7 Food and Feeding

10.3.7.1 This section is based on recommendations made by Rhoads et al. (1972) and Cripe (2006). Laboratories unfamiliar with this species should utilize the specific diet recommended below. Modifications to the diet could then proceed by laboratories in order to optimize culture practices as long as the modifications satisfied the performance criteria.

10.3.7.2 Holding chambers should be provided with food in conjunction with water renewal. On a daily basis, approximately 10% of the culture water should be removed from each culture chamber and replaced with the same volume of phytoplankton. Each tank should be fed a volume of 5% of the holding chamber of each of *Isochrysis* sp. (T-ISO) and *Tetraselmis striata,* with addition of *Chaetoceros calcitrans,* if available. The cultured phytoplankton should be oligoxenic monocultures, and cell density used for feeding adult clams should approach 2 to 3 × 10^6 *Isochrysis* cells/mL and 1 to 2 × 10^5 *Tetraselmis* cells/mL. These may be fed to holding tanks up to a 50/50 combination. Deionized water should be added to each culture to maintain the salinity, with resultant changes of not greater than 3‰ per day. Other algal species can be used if it can be demonstrated that they foster clam growth and reproductive rates equal to those of the aforementioned algal species.

10.3.8 Maintenance

10.3.8.1 Observations and Measurements. Holding tanks should be observed daily to ensure that aeration is adequate in all culture chambers. Inspection for the presence of competing or predacious organisms should be conducted weekly. The presence of excessive densities of these organisms should prompt renewal of culture sediment after removing the clams from the sediment.

10.3.8.2 Healthy populations are characterized by an abundance of extended siphons at the sediment/water interface. If the sediment becomes an unacceptable habitat (i.e., if it is black and sulfurous below the sediment surface) or contains an excess of competitive or predacious organisms, the healthy surviving clams should be placed in a new holding tank with clean sediment.

10.3.8.3 Water temperature and salinity should be measured daily. Cultures should be continuously aerated.

10.4 Algal Culture

10.4.1 Algae. Three algal species are used to culture and maintain *M. lateralis* broodstock, larvae, and juveniles: *Isochrysis* sp. (T-ISO) *(Iso),Chaetoceros calcitrans (Cha),* and *Tetraselmis striata (Tetra). Iso* and *Cha* are used to culture *M. lateralis* larvae. *M. lateralis* juveniles and broodstock are fed with *Iso* and *Tetra.* Axenic mono cultures of algae can be obtained from Milford NMFS laboratory in Milford, CT, or from commercial suppliers. These are shipped as test tube starter cultures. All algal species are cultured between 25‰ and 30‰.

10.4.2 The test tube starter cultures should be transferred aseptically into 500-mL bottles containing autoclaved seawater that has been filtered to 1 µm. The bottles should also contain F/2 media (Guillard and Ryther 1962). The cultures can be maintained axenic by transferring 5 mL of culture into a new 500-mL bottle containing filtered, autoclaved seawater and F/2 nutrients. The bottle should be aerated with a sterile 1-mL pipet and stoppered with a foam plug. The cultures should be maintained at 21 °C ±2 °C on a light:dark cycle (16 h light: 8 h dark or 12 h light: 12 h dark). Lighting for the cultures can be provided by fluorescent lights 10 cm away from the cultures with a recommended light intensity of 9000 to10,000 lux. These small bottle cultures are used to start larger carboy cultures as well as to feed juvenile tests.

10.4.3 Larger carboy cultures are used to feed the broodstock as well as the larvae and juvenile cultures. A 500-mL starter culture should be aseptically transferred to a 20-L carboy containing autoclaved seawater filtered to 1 µm and F/2 nutrients. The carboy should be aerated constantly with a sterile 10 mL pipet and stoppered with an autoclaved foam plug. The carboy cultures should be maintained with the same lighting at the same photoperiod as the starter cultures. Alternatively, large cultures may be maintained in disposable food grade plastic bags containing seawater filtered through a 0.45-µm hydrophilic cartridge filter. When using this technique with *Chaetoceros,* sodium metasilicate must be added to the water as a silica source.

10.5 Spawning

10.5.1 Adult male and female *M. lateralis* should be induced to spawn by temperature shock (more detail on the method can be found in Petrocelli and Morrison 1990). Adults should be rinsed in clean seawater and placed in bowls containing clean seawater (either individually separated in small cups if sex is unknown or separated by sex in larger bowls). The water can then be cooled to 4 °C by placing the bowls or cups in a refrigerator or by placing them over ice for 1 to 2 h. After cooling, the animals are rinsed with clean seawater and placed in their bowls or cups with clean seawater at ambient temperature. The water is then warmed to 25 to 28 °C by placing the bowls or cups in an incubator or by placing them in a warm water baths. Spawning temperature should not exceed 29 °C, or damage to the embryos could result (Kennedy et al. 1974). The animals should begin to spawn shortly after being warmed, although sometimes several hours may pass before spawning is observed. If no spawning has occurred within 2 h of warming, the animals should be returned to their holding tanks and spawning should be attempted with a different batch of animals or at another time. Sperm may be identified by a white, milky appearance, while eggs will appear pinkish or orange and of a more granular texture. Spawning males should be removed from the water after a few seconds (before the spawning water becomes cloudy) and placed into another dish containing clean seawater. After collection of the gametes, the eggs and sperm should be combined at a ratio of about 300 sperm/egg (Sheehan and Pelletier 1992). After fertilization, the embryos should be added to vessels with completely flat bottoms, as the larvae are poor swimmers and passively accumulate in depressions. This results in degradation of local water conditions and high

mortality. Larvae may be cultured at 12 to 30 per mL of culture water. These culture chambers should contain natural seawater filtered to 0.45 µm and aerated very gently using a pipet.

10.6 Larval Culture

10.6.1 Larval cultures should be maintained at 24 ±1 °C, 500-1000 lux and culture water should be renewed every other day by very slowly passing the water containing larvae through a sieve constructed from nylon mesh that is partially submerged in seawater to minimize impingement of larvae on the screen. For example, a 36-µm screen should be used for renewals on days 2, 4 and 6, then a 55-µm screen should be used for renewals thereafter. Seawater filtered to 0.45 µm or sterilized must be used for culture prior to metamorphosis. At the time of renewal, each culture chamber should be supplied a mixture of *Isochrysis* sp. (T-ISO) and *Chaetoceros calcitrans.* Maintenance of an algal concentration of no less than 1.0×10^5 cells/mL in each culture vessel should provide adequate food to the larvae. Note that each larva will clear about 200 cells/day between days 2 and 4 and this will increase to 700 cells/day by day 8. This rate will be maintained until metamorphosis at day 14. If necessary, the salinity of the seawater/algae mix should be adjusted to 27‰ to 30‰. The cultures should be moderately aerated at all times to keep the larvae distributed in the water column.

10.7 Juvenile Culture

10.7.1 Fourteen days after fertilization, most of the larvae should have metamorphosed (lost their velum and developed siphons). Clams may be grown to testing size in either static or semi static systems. Once the larvae have metamorphosed they can be placed, at a density of 300/L, in downwellers (Hadley et al. 1997) suspended in culture aquaria containing 1-µm filtered or autoclaved seawater. Downwellers are constructed by gluing 100-µm nylon mesh with silicone cement to the bottom of a plastic cylinder 10-cm high, 15-cm diameter (Cripe 2006). The upper side of the cylinder contains an air lift tube that functions to move the water into the sieve, resulting in a mass flow of algae-laden water downward over the clams. Maintaining clams on these sieves facilitates transfer to new culture water. As the juveniles grow, transfer to downwellers with 500-µm screen will facilitate cleaning. At this time they can be fed a mixture of *Isochrysis* sp. (T-ISO) and *Tetraselmis striata.* After metamorphosis to settled stage, (day 14) algal-clearance rate increases, and each juvenile may clear up to 6000 cells per day by testing size. Water should be renewed every other day in static systems. The culture chambers should be brushed and rinsed with deionized water before refilling with clean seawater and algae. The juveniles can be gently rinsed with a squirt bottle before returning to the culture chamber. Rinsing will minimize the buildup of dead algae clumps on the shells of the juveniles. Semistatic culture systems (Cripe 2006) may be maintained with daily infusions of clean seawater and algal cultures, with excess water removed by overflow to a tank drain, with tank cleaning every 5 to 7 days. Survival of 50% of juvenile clams from metamorphosis to testing size is acceptable.

Section 11

Mulinia lateralis 10-Day Survival and Growth Test for Whole Sediments

11.1 Introduction

11.1.1 *Mulinia lateralis* has been used to test the toxicity of estuarine and marine sediments (Burgess and Morrison 1994, Burgess et al. 1994). The choice of this species as a test organism is based on its sensitivity to sediment-associated contaminants, availability and ease of collection, tolerance to environmental conditions (e.g., temperature, salinity, grain-size), ecological importance, and ease of handling in the laboratory. Additionally, this species is intimately associated with sediment by nature of its burrowing and feeding habits. Thus, it may be used to measure toxicity associated with commonly encountered estuarine or marine sediment.

11.1.2 Specific test methods for conducting the 10-day sediment toxicity test with the bivalve *M. lateralis* are described in section 11.2. This test method was developed based on Burgess and Morrison (1994). Results of tests using procedures different from procedures described in section 11.2 may not be comparable and may alter contaminant bioavailability. Comparison of results obtained using modified versions of these procedures might provide useful information concerning new concepts and procedures for conducting sediment tests with estuarine or marine organisms. However, if tests are conducted with procedures different from those described in the manual, additional tests will be required to determine comparability of results (section 1.3).

11.2 Recommended Test Method for Conducting a 10-Day Sediment Toxicity Test with *Mulinia lateralis*

11.2.1 Recommended conditions for conducting a 10-day sediment toxicity test with *M. lateralis* are summarized in Table 11.1. A general activity schedule is outlined in Table 11.2. Decisions concerning the various aspects of experimental design, such as the number of treatments, number of replicates, and water quality characteristics should be based on the purpose of the test and the methods of data analysis (section 12). When variability remains constant, the sensitivity of a test increases as the number of replicates increases. The number of replicates and concentrations tested will depend, in part, on the significance level selected and the type of statistical analysis. Five replicates are recommended for routine testing.

11.2.2 The recommended 10-day sediment toxicity test with *M. lateralis* must be conducted at 22 °C and 30‰ under constant light at an illuminance of approximately 2000-4000 lux (Table 11.1). Test chambers are 250-mL glass chambers containing approximately 50 mL of sediment and 150 mL of overlying seawater. Ten clams are added to each test chamber at the start of a test. Clams of the appropriate size for testing must pass through a 1.5-mm sieve and be retained on a 1.0-mm sieve. Exposure is static (i.e., water is not renewed), and the animals fed daily a 50/50 combination diet of *Isochrysis* sp. (T-ISO) and *Tetraselmis striata* resulting in 1×10^5 cells/mL of overlying water. Overlying water can be culture water, surface water, site water, or reconstituted water. For site-specific evaluations, the characteristics of the overlying water should be as similar as possible to the site where sediment is collected. For all other applications, the characteristics of the overlying water for each species should be chosen according to Table 11.1. Requirements for test acceptability are summarized in Table 11.3.

Table 11.1 RECOMMENDED TEST CONDITIONS FORCONDUCTING A 10-DAY SEDIMENT TOXICITY TEST WITH *MULINIA LATERALIS*

Parameter		Conditions
1.	Test type:	Static, 10-day duration
2.	Species:	*Mulinia lateralis*
3.	Test chamber	250-mL glass jar, internal diameter approximately 10 cm
4.	Control sediment:	50 mL clean sediment, may be from clam collection site; sieved through 0.5 mm screen
5.	Test water:	150 mL clean seawater, natural or reconstituted; placed above the layer of sediment in test chamber the day before the test is initiated, then renewed the day of test initiation
6.	Temperature:	22 $\pm$2 °C
7.	Salinity:	30 $\pm$2‰ (tolerant of 7‰ to >34‰)
8.	Aeration:	None, unless pH or dissolved oxygen is below criteria
9.	Lighting:	Constant; 2000-4000 lux
10.	Size and life stage of clams:	Juveniles, 1 to 1.5 mm
11.	Number of clams per test chamber:	10
12.	Number of replicates:	Depends on objectives of test; At a minimum, three replicates must be used and five are recommended.
13.	Feeding:	Daily, combination of *Isochrysis* sp. (T-ISO) and *Tetraselmis* striata to result in 1×10^5 cells per mL of overlying water
14.	Observations:	Daily, each test vessel, to observe the number of dead clams on sediment surface
15.	Measurements:	Temperature and salinity daily; pH, ammonia, and DO of overlying water at least at test start and end; salinity, ammonia, and pH of pore water. All these should be measured on one replicate daily and on all chambers at start and end of a test
16.	Test duration:	10-days

Table 11.1 RECOMMENDED TEST CONDITIONS FORCONDUCTING A 10-DAY SEDIMENT TOXICITY TEST WITH *MULINIA LATERALIS (cont'd.)*

Parameter		Conditions
17.	End points:	Survival and growth (dry weight)
18.	Test acceptability:	Minimum mean control survival of 80%, controls must double in weight, and satisfaction of performance-based criteria specifications outlined in Table 11.3 is required.

Table 11.2 GENERAL ACTIVITY SCHEDULE FOR CONDUCTING A SEDIMENT TOXICITY TEST WITH MULINIA LATERALIS

Day	Activity
-10 to -3	Collect or receive clams from supplier and place into noncontaminated sediment.
-9 to -2	Acclimate clams to test conditions, feed, monitor water quality (e.g., temperature, salinity, dissolved oxygen) and observe behavior.
-1	Observe clams, monitor water quality. Add sediment to each test chamber, place chambers into exposure system.
0	Measure pore water total ammonia, salinity, and pH. Measure temperature of overlying water in test chambers. Renew overlying water in test chambers. Transfer 10 clams into each test chamber. Place necessary number of replicates with 10 animals each into drying oven for initial weight determination. Feed test.
1	Measure temperature. Observe behavior of test organisms. Feed test.
1	Measure total water quality (pH, temperature, dissolved oxygen, salinity, and total ammonia of overlying water). Observe behavior of test organisms and feed.
3 to 7 and 9	Same as Day 1.
8	Same as Day 2.
10	Measure temperature. End the test by collecting the clams with a sieve. Determine survival, then place live animals into drying oven for weight determination.
12 to 14	Weigh dried clams for weight determination.

Table 11.3 TEST ACCEPTABILITY REQUIREMENTS FOR A 10-DAY SEDIMENT TOXICITY TEST WITH *MULINIA LATERALIS*

A. It is recommended for conducting a 10-day test with *M. lateralis* that the following performance criteria are met.

1. Average survival of clams in the control sediment must be greater than or equal to 80% at the end of the test.

2. Controls must double in weight.

3. Salinity, pH, and ammonia in the overlying water are within tolerance limits of *M. lateralis.*

4. DO must not fall below 4.0 mg/L.

B. Performance-based criteria for culturing *M. lateralis*

1. Laboratories should perform monthly 96 h water-only reference-toxicity tests to assess the sensitivity of culture organisms. If reference-toxicity tests are not conducted monthly, the lot of organisms used to start a sediment test must be evaluated using a reference toxicant (section 9.16).

2. Records should be kept on the frequency of restarting cultures.

3. Laboratories should record the pH and ammonia concentrations of the cultures at least quarterly. Salinity should be measured weekly. Temperature should be recorded daily.

4. Laboratories should characterize and monitor background contamination and nutrient quality of food if problems are observed in culturing or testing organisms.

C. Performance-based criteria for field-collected *M. lateralis*

1. Laboratories should perform reference-toxicant tests on each batch of field-collected clams used in a sediment test (section 9.16).

2. Acclimation rates to test salinity and temperature should not exceed 3 °C and 3‰ per 24-h.

3. Clams received from commercial suppliers must exhibit active behavior upon placement in water, have full digestive tracts, and have an acceptable coloration.

Table 11.3 TEST ACCEPTABILITY REQUIREMENTS FOR A 10-DAY SEDIMENT TOXICITY TEST WITH *MULINIA LATERALIS* *(cont'd.)*

D. Additional requirements

1. All organisms in a test must be from the same source.
2. It is desirable to start tests as soon as possible after collection of sediment from the field (see Section 8.2 for additional detail).
3. All test chambers (and compartments) should be identical and should contain the same amount of sediment and overlying water.
4. Negative-control sediment must be included in a test.
5. The time-weighted average of daily temperature readings must be within ±2 °C of the desired temperature. The instantaneous temperature must always be within ±3 °C of the desired temperature.

Natural physico-chemical characteristics of test sediment collected from the field should be within the tolerance limits of the test organism.

11.2.3 If the recommended test temperature and/or salinity are not appropriate for some localities (i.e. estuarine sediments or Gulf of Mexico temperature regimes) these may be adjusted as long as they fall within the tolerance range of the species. Please be aware that contaminant bioavailability (see chapter 4) may be affected when temperature and/or salinity is modified. The results of tests with such manipulations may not be comparable to those performed under standard conditions.

11.3 General Procedures

11.3.1 Introduction of Sediment. On the day before the addition of clams (Day-1), each test sediment (either field collected or laboratory spiked) should be homogenized by stirring in the sediment storage container or by using a rolling mill, feed mixer, electric drill, or other suitable apparatus. Control and reference sediments should be treated similarly. Sediment should be visually inspected to judge the extent of homogeneity. Excess water on the surface of the sediment can indicate separation of solid and liquid components and must be homogenized before use. If a quantitative measure of homogeneity is required, replicate subsamples should be taken from the sediment batch and analyzed for TOC, chemical concentrations, and particle size. Spiked sediments should not be homogenized until the overlying water has been drained off because the equilibrium between the spiked contaminant and the sediment partitioning factors may be disrupted.

11.3.1.1 A 50-mL aliquot of thoroughly homogenized sediment is added to each test chamber. It is important that an identical volume be added to each replicate test chamber. The sediment added to the test chamber should be evenly distributed either by tapping the side of the test chamber against the side of the hand or by smoothing the sediment surface with a nylon,

fluorocarbon, or polyethylene spatula. Highly contaminated sediment should be added to test chambers in a certified laboratory fume hood.

11.3.2 Addition of Overlying Water. To minimize disruption of sediment as test seawater is added, a turbulence reducer should be used. The turbulence reducer may be either a disk cut from polyethylene, nylon, or Teflon sheeting (4 to 6 mil) or a glass Petri dish attached (open face up) to a glass pipet. If a disk is used as the turbulence reducer, it should fit the inside diameter of the test chamber and have attached a length of nylon monofilament (or nontoxic equivalent) line. The turbulence reducer is positioned just above the sediment surface and raised as seawater is added to the 200-mL mark on the side of the test chamber. The turbulence reducer is removed and rinsed with test seawater between replicates of a treatment. A separate turbulence reducer should be used for each treatment. The test chambers should be covered, placed in a temperature controlled water bath (or other acceptable equivalent). The following day (Day 0), the overlying water is gently poured off and replaced with fresh overlying water to remove hydrogen sulfide and ammonia that might have been introduced from the sediment to the water during exposure preparation. The test begins when the organisms are added to the test chambers (Day 0).

11.3.3.1 Addition of clams. Clams are removed from the culture tank and sorted to the correct size by sieving. Approximately one-third more clams than are needed for the test should be initially collected to allow for the selection of healthy, active individuals. Sieving should be conducted with seawater of the same temperature and salinity as the holding and test water. Once isolated and transferred to a sorting dish, healthy clams should be randomly selected using a transfer pipet or other suitable tool (not forceps), and distributed among 20-mL cups containing approximately 15 mL of test seawater until each container has ten clams. Healthy clams are pinkish in coloration with a dark, full gut. The number of clams in each dish should be verified by recounting before adding to test chambers. The distribution of clams to the test chambers must be executed in a randomized fashion. An additional "treatment" containing the same number of individuals and replicates should also be prepared in order to obtain the average initial weights of the clams. These clams should be prepared by rinsing the clams twice with deionized water in a small Petri dish, eyecup, or beaker. Remove seawater using a pipet and replace with deionized water. Replace and drain the deionized water a second time. Transfer the clams using a wide bore pipet to a preweighed weight boat with minimal transfer of water. Pipet off excess water before placing in a drying oven at 60 °C for 48 h prior to weight determination.

11.3.3.2 Clams should be added to test chambers without disruption of the sediment by gently washing them from their sorting cups into the test chamber with test seawater. The test chambers should be covered to prevent evaporation.

11.3.3.3 After the addition of the animals, the test chambers should be examined for animals that may have been injured or stressed during the isolation, counting, or addition procedures. *M. lateralis* should be allowed a half-hour to bury into test sediments. If the clams are moving on the sediment or have extended their siphons to feed, they should be considered healthy and not be removed from the test. Those animals that are not active or burrowing should be removed and replaced with healthy animals from the same sieved population.

11.3.4 Test Conditions

11.3.4.1 Aeration. The overlying seawater in each test chamber is not aerated unless it is needed to stabilize pH or to increase dissolved oxygen. If aeration is necessary, compressed

air, previously filtered and free of oil, should be bubbled through a glass or plastic pipet and attached plastic tubing. If it is necessary to aerate any treatment, all must be aerated. The tip of the pipet should be suspended 2 to 3 cm above the surface of the sediment layer to prevent disturbing the sediment surface. If not necessary, aeration should be avoided to minimize evaporation. Usually an acceptable concentration of dissolved oxygen (DO) in the water overlying the sediment in the test chambers is maintained by the photosynthesizing phytoplankton. Although *M. lateralis* can tolerate dissolved oxygen levels as low as 2.0 mg/L without any growth or survival effects (Rego, unpublished data), DO should remain above 4.0 mg/L throughout the test.

11.3.4.2 Lighting. Lights must be left on continuously at an intensity of 2000-4000 lux during the 10-day exposure period. Constant lighting results in a net production of oxygen by the algae through photosynthesis and prevents dissolved oxygen levels from declining below acceptable limits.

11.3.4.3 Feeding. *Mulinia lateralis* must be fed a combination of *Isochrysis galbana* and *Tetraselmis striata* in sufficient amounts to maintain a concentration of 1×10^5 cells of algae per mL of overlying water within the testing chambers daily during the 10-day exposure period. This food concentration will ensure that food is available in excess and will not need to be adjusted to the number of clams when some have died.

11.3.4.4 Water Temperature. The test must be conducted at 22 ±2 °C .

11.3.4.5 Salinity. For routine testing, *M. lateralis* should be tested at an overlying water salinity of 30‰. The pore water salinity of the test sediment must be within the salinity application range of *M. lateralis* (30 ±2‰: tolerates 7 to >34‰; Table 11.1). Sediment pore water should be obtained by centrifugation. Alternatively, salinity can be measured before homogenization in the water that comes to the surface in the sample container as the sediment settles. Salinity should be measured during the test and adjusted with deionized water when necessary.

11.3.5 Measurements and Observations

11.3.5.1 Temperature should be measured at least daily in at least one test chamber from each treatment. The temperature of the water bath or the exposure chamber should be continuously monitored. The time-weighted average of daily temperature readings must be within ±2 °C of the desired temperature. The instantaneous temperature must always be within ±3 °C of the desired temperature.

11.3.5.2 Dissolved oxygen and pH of the overlying water should be measured daily in at least one test chamber per treatment, and in every test chamber at the beginning (e.g., day 2) and the end (e.g., day 8) of a test.

11.3.5.3 Ammonia should be measured in overlying water towards the beginning (e.g., day 2) and towards the end of the test (e.g., day 8) in each test chamber. Measurement of overlying water pH, salinity, and temperature should accompany each ammonia measurement. Simultaneous measurements of ammonia, salinity, pH, and temperature in sediment pore water should be measured at the beginning of the test. Pore water should be extracted after the sediment has been press-sieved and homogenized. Samples of pore water should be obtained by centrifugation.

11.3.5.4 Each test chamber must be examined at least daily during the 10-day test period. The number of dead or moribund animals should be noted.

11.3.6 Ending a Test. Laboratories should demonstrate the ability of their personnel to recover an average of at least 80% of the organisms from control sediment. For example, prior to conducting a test, extra organisms could be added to control sediment and recovery attempted after 2 h to determine total recovery (Tomasovic et al. 1994).

11.3.6.1 Contents of test chambers must be sieved to isolate the test organisms. The mesh size for sieving the contents of the test chambers must be no larger than 1.0 mm. Test water should be used for sieving. Material retained on the sieve should be washed into a sorting dish with clean test seawater.

11.3.6.2 Material that has been washed from the sieve into the sorting dish should be carefully examined for the presence of clams. A small portion of the material should be sorted through at a time, removing clams as they are found. Numbers of live, missing, and dead clams should be determined and recorded for each test chamber. Missing animals are assumed to have died, decomposed, and disintegrated during the test and they should be included in the number dead in calculations of the percent survival for each replicate treatment. Clams that are inactive but not obviously dead must be observed using a low-power dissecting microscope or compound microscope. Those animals not actively moving their foot or siphons should be examined under a microscope for a heartbeat or gill movement. If no movement can be discerned, a small amount of algae should be added to the dish. If siphoning or other movement does not occur within 3 to 5 min, the animal should be considered dead. The live animals from each replicate must be rinsed with deionized water in the same manner as in section 11.3.3.1.

11.3.7 Test Data. Percent survival is the number of live animals recovered at the end of the test divided by the number of animals added to the test. Growth, assessed as the final weight of live clams, is determined by dividing the total weight by the number of live clams recovered.

11.4 Interpretation of Results

11.4.1 Section 12 describes general information for interpretation of test results.

11.4.2. Influence of Indigenous Organisms. Indigenous organisms may be present in field-collected sediments. In the sediment sample, an abundance of the test organism, or organisms taxonomically similar to the test organism, may make interpretation of treatment effects difficult. Presence of predatory organisms can also adversely affect test organism survival. For example, Luchenbach (1984) showed that polychaetes will reduce *Mulinia lateralis* survival under test conditions.

11.4.3. Effect of Sediment Grain Size. *Mulinia lateralis* shows tolerance to most sediment types, with generally little effect on survival or growth whether coarse-grained or fine-grained (i.e., predominantly silt and clay) clean sediments are used. However, adverse effects due to the grain-size distribution of test sediment may occur when sediments are either extremely sandy or fine. In order to separate effects of sediment-associated contaminants from effects of particle size, an appropriate clean control/reference sediment should be incorporated into the test when test sediment grain size is of concern.

11.4.4 Effects of Pore Water Salinity. *Mulinia lateralis* is euryhaline (7% to >34%). Therefore, either the standard salinity for the test species can be used or the salinity can be matched to that

of the pore water. If a test salinity other than 30‰ is used, a 30‰ control treatment should be incorporated in order to determine salinity effects.

11.4.5 Effects of Sediment-Associated Ammonia. Field-collected sediments may contain concentrations of ammonia that are toxic to juvenile clams. Water column LC50s for survival of *M. lateralis* are 21.7 mg/L total ammonia and 0.6 mg/L unionized ammonia (Table 11.4). Water column EC50s for growth of *M. lateralis* are 11.0 mg/L total ammonia and 0.3 mg/L unionized ammonia (Table 11.4). If ammonia concentrations are above the LC50 values, mortality occurring after 10-days may be due in part to effects of ammonia. Depending on the test application, it may be desirable to lower the ammonia concentration by manipulating the test system prior to introducing test organisms if measured ammonia in the overlying water is greater than the levels cited above. If sediment toxicity tests are conducted to evaluate the acceptability of dredge material for disposal, these manipulations must be performed. Manipulations involve flushing the test system by renewing a specified amount of overlying water for up to two consecutive 24-h periods.

11.4.5.1 If ammonia is of concern to the regulatory application associated with the sediment toxicity test, overlying water should be sampled approximately 1 cm above the sediment surface prior to introduction of animals on day 0. If overlying water total ammonia concentration is less than or equal to 21.7 mg/L, then the test may proceed normally. Otherwise the test system must be flushed for 24 h at a rate of six volume replacements per 24 h.

11.4.5.2 After 24 h, the overlying water ammonia concentration must be measured again. If it is less than or equal to 21.7 mg/L total ammonia, testing should be initiated by adding animals. The system must be flushed at a rate of six volume replacements per 24 h over the course of the test. Overlying water ammonia should be measured again on day 10 of the test.

11.4.5.3 If after the initial 24 h flushing period (i.e., that described in section 11.4.5.1) the overlying water ammonia concentration is still greater than 21.7 mg/L total ammonia, the system must be flushed for again 24 h at a rate of six volume replacements per 24 h. After the second flushing, ammonia concentrations in the overlying water should be measured again, and if concentrations are less than or equal to 21.7 mg/L total ammonia, then the test may proceed as described in section 11.4.5.2. If overlying water ammonia concentration still exceed the species-specific no effect concentration, it must be concluded that ammonia cannot be reduced to a no-effect concentration without concern for flushing other contaminants from the sediment. At this point, the test should still be conducted as described in section 11.4.5.2. Additionally, sediment chemical analyses and bioaccumulation tests should be conducted as recommended by USEPA-USCOE (1991). Results of all required Tier III tests should be evaluated simultaneously following guidance in USEPA-USCOE (1991).

Table 11.4 APPLICATION LIMITS FOR 10-DAY SEDIMENT TOXICITY TESTS WITH *MULINIA LATERALIS*

Parameter	*Mulinia lateralis*
Temperature (°C)	20 to 24
Overlying Salinity (l)	7 to >32
Grain Size (% silt/clay)	Full range
Dissolved Oxygen (mg/L)	>4.0
SURVIVAL	
Ammonia (total mg/L)	>21.7
Ammonia (UI [a] mg/L)	<0.6
GROWTH	
Ammonia (total mg/L)	<11.0
Ammonia (UI [a] mg/L)	<0.3

[a] UI = unionized ammonia

Section 12

Data Recording, Data Analysis and Calculations, and Reporting

12.1 Data recording

12.1.1 Quality Assurance Project Plans with Data Quality Objectives and Standard Operating Procedures should be developed before starting a test. Procedures should be developed by each laboratory to verify and archive data.

12.1.2 A file should be maintained for each sediment test or group of tests on closely related samples (section 9). This file should contain a record of the sample chain-of-custody; a copy of the sample log sheet; the original bench sheets for the test organism responses during the sediment test(s); chemical analysis data on the sample(s); control data sheets for reference toxicants; detailed records of the test organisms used in the test(s), such as species, source, age, date of receipt, and other pertinent information relating to their history and health; information on the calibration of equipment and instruments; test conditions used; and results of reference toxicant tests. Original data sheets should be signed and dated by the laboratory personnel performing the tests.

12.1.3 Example data sheets are included in Appendix A.

12.2 Data Analysis

12.2.1 Statistical methods are used to make inferences about populations, based on samples from those populations. In most sediment toxicity tests, test organisms are exposed to contaminated sediment to estimate the response of the population of laboratory organisms. The organism response to these contaminated sediments is usually compared with the response to a control or reference sediment. In any toxicity, summary statistics such as means and standard errors for response variables (e.g., survival, growth) should be provided for each treatment (e.g., pore-water concentration, sediment).

12.2.1.1 Types of data. Two types of data can be obtained from sediment toxicity tests. The most common end point in toxicity testing is mortality, which is a dichotomous or categorical type of data.

12.2.1.2 Sediment Testing Scenarios. Sediment tests are conducted to determine whether contaminants in sediment are harmful to or are bioaccumulated in benthic organisms. Sediment tests are commonly used in studies designed to (1) evaluate hazards of dredged material, (2) assess site contamination in the environment (e.g., to rank areas for clean-up), and (3) determine effects of specific contaminants or combinations of contaminants through the use of sediment-spiking techniques. Each of these broad study designs has specific statistical design and analytical considerations, which are detailed below.

12.2.1.2.1 Dredged Material Hazard Evaluation. In these studies, “n” sites are compared individually to a reference sediment. The statistical procedures appropriate for these studies are generally pairwise comparisons. Additional information on toxicity testing of dredged material and analysis of data from dredged material hazard evaluations is available in USEPA-USCOE (1994).

12.2.1.2.2 Site Assessment of Field Contamination. Surveys of sediment toxicity often are included in more comprehensive analyses of biological, chemical, geological, and hydrographic data. Statistical correlation can be improved and costs may be reduced if subsamples are taken simultaneously for sediment toxicity, chemical analyses, and benthic community structure determinations. There are several statistical approaches to field assessments, each with a specific purpose. If the objective is to compare the response or residue level at all sites individually to a control sediment, then the pairwise comparison approach described below is appropriate. If the objective is to compare among all sites in the study area, then a multiple comparison procedure that employs an experiment-wise error rate is appropriate. If the objective is to compare among groups of sites, then orthogonal contrasts are a useful data analysis technique.

12.2.1.2.3 Sediment Spiking Experiments. Sediments spiked with known concentrations of contaminants can be used to establish cause and effect relationships between chemicals and biological responses. Results of toxicity tests with test materials spiked into sediments at different concentrations may be reported in terms of an LC50, EC50, IC50, NOEC, or LOEC. The statistical approach outlined above for spiked sediment toxicity tests also applies to the analysis of data from sediment dilution experiments or water-only reference toxicant tests.

12.2.2 The guidance outlined below on the analysis of sediment toxicity test data is adapted from a variety of sources including Lee et al. (1989), USEPA (1993a,b,c), and USEPA-USCOE (1994). The objectives of a sediment toxicity test is to quantify contaminant effects on test organisms exposed to natural or spiked sediments or dredged materials and to determine whether these effects are statistically different from those occurring in a control or reference sediment. Each experiment consists of at least two treatments: the control and one or more test treatment(s). The test treatment(s) consist(s) of the contaminated or potentially contaminated sediment(s). A control sediment is always required to ensure that no contamination is introduced during the experimental setup and that test organisms are healthy. A control sediment is used to judge the acceptability of the test. Some designs will also require a reference sediment that represents an environmental condition or potential treatment effect of interest.

12.2.2.1 Experimental Unit. During toxicity testing, each test chamber to which a single application of treatment is applied is an experimental unit. The important concept is that the treatment (sediment) is applied to each experimental unit as a discrete unit. Experimental units should be independent and should not differ systematically.

12.2.2.2 Replication. Replication is the assignment of more than one experimental unit to a treatment. The variation among replicates is a measure of the within-treatment variation and provides an estimate of within-treatment error for assessing the significance of observed differences between treatments.

12.2.2.3 Minimum Detectable Difference (MDD). As the minimum difference between treatments that the test is required or designed to detect decreases, the number of replicates required to meet a given significance level and power increases. Because no consensus currently exists on what constitutes a biologically acceptable MDD, the appropriate statistical minimum significant difference should be a data quality objective (DQO) established by the individual user (e.g., program considerations) based on their data requirements, the logistics and economics of test design, and the ultimate use of the sediment toxicity test results.

12.2.2.4 Minimum number of replicates. Three replicates per treatment or control are the absolute minimum number of replicates for a sediment toxicity test. However, USEPA recommends five replicates for marine testing or eight replicates for freshwater testing (USEPA 1994a) for each control or experimental treatment. It is always prudent to include as many replicates in the test design as are economically and logistically possible. USEPA sediment toxicity testing methods recommend the use of 10 organisms per replicates for marine or freshwater testing (USEPA 1994a). An increase in the number of organisms per replicate in all treatments, including the control, is allowable only if (1) test performance criteria for the recommended number of replicates are achieved and (2) it can be demonstrated that no change occurs in contaminant availability due to the increased organism loading.

12.2.2.5 Randomization. Randomization is the unbiased assignment of treatments within a test system and to the exposure chambers ensuring that no treatment is favored and that observations are independent. It is also important to (1) randomly select the organisms (but not the number of organisms) for assignment to the control and test treatments (e.g., a bias in the results may occur if all the largest animals are placed in the same treatment), (2) randomize the allocation of sediment (e.g., not take all the sediment in the top of a jar for the control and the bottom for spiking), and (3) randomize the location of exposure units.

12.2.2.6 Pseudoreplication. The appropriate assignment of treatments to the replicate exposure chambers is critical to the avoidance of a common error in design and analysis termed "pseudoreplication" (Hurlbert 1984). Pseudoreplication occurs when inferential statistics are used to test for treatment effects even though the treatments are not replicated or the replicates are not statistically independent (Hurlbert 1984). The simplest form of pseudoreplication is the treatment of subsamples of the experimental unit as true replicates. For example, two aquaria are prepared, one with control sediment, the other with test sediment, and 10 organisms are placed in each aquarium. Even if each organism is analyzed individually, the 10 organisms only replicate the biological response and do not replicate the treatment (i.e., sediment type). In this case, the experimental unit is the 10 organisms and each organism is a subsample. A less obvious form of pseudoreplication is the potential systematic error due to the physical segregation of exposure chambers by treatment. For example, if all the control exposure chambers are placed in one area of a room and all the test exposure chambers are in another, spatial effects (e.g., different lighting, temperature) could bias the results for one set of treatments. Random physical intermixing of the exposure chambers or randomization of treatment location may be necessary to avoid this type of pseudoreplication. Pseudoreplication can be avoided or reduced by properly identifying the experimental unit, providing replicate experimental units for each treatment, and applying the treatments to each experimental unit in a manner that includes random physical intermixing (interspersion) and independence. However, avoiding pseudoreplication completely may be difficult or impossible given resource constraints.

12.2.3 The purpose of a toxicity test is to determine if the biological response to a treatment sample differs from the response to a control sample. Table 12.1 presents the possible outcomes and decisions that can be reached in a statistical test of such a hypothesis. The null hypothesis is that no difference exists among the mean control and treatment responses. The alternative hypothesis of greatest interest in sediment tests is that the treatments are toxic, or contain concentrations of bioaccumulable compounds, relative to the control or reference sediment.

Table 12.1 SUGGESTED α LEVELS TO USE FOR TESTS OF ASSUMPTIONS

Test	Number of Observations [a]	α When Design Is	
		Balanced	Unbalanced [b]
Normality	N = 2 to 9	0.10	0.25
	N = 10 to 19	0.05	0.10
	N = 20 or more	0.01	0.05
Equality of Variances	n = 2 to 9	0.10	0.25
	n = 10 or more	0.05	0.10

[a] N = total number of observations (replicates) in all treatments combined; n = number of observations (replicates) in an individual treatment

[b] $n_{max} \geq 2n_{max}$

12.2.3.1 Statistical tests of hypotheses can be designed to control for the chances of making incorrect decisions. In Table 12.1, alpha (α) represents the probability of making a Type I statistical error. A Type I statistical error in this testing situation results from the false conclusion that the treated sample is toxic or contains chemical residues not found in the control or reference sample. Beta (β) represents the probability of making a Type II statistical error, or the likelihood that one erroneously concludes there are no differences among the mean responses in the treatment, control or reference samples. Traditionally, acceptable values for α have ranged from 0.1 to 0.01 with 0.05 or 5% used most commonly. This choice should depend upon the consequences of making a Type I error. Historically, having chosen α, environmental researchers have ignored β and the associated power of the test (1-β).

12.2.3.2 Fairweather (1991) presents a review of the need for, and the practical implications of, conducting power analysis in environmental monitoring studies. This review also includes a comprehensive bibliography of recent publications on the need for, and use of, power analyses in environmental study design and data analysis. The consequences of a Type II statistical error in environmental studies should never be ignored and may in fact be the most important criteria to consider in experimental designs and data analyses which include statistical hypothesis testing. To paraphrase Fairweather (1991), "The commitment of time, energy and people to a false positive (a Type I error) will only continue until the mistake is discovered. In contrast, the cost of a false negative (a Type II error) will have both short- and long-term costs (e.g., ensuing environmental degradation and the eventual cost of its rectification)."

12.2.3.3 The critical components of the experimental design associated with the test of hypothesis outlined above are (1) the required MDD between the treatment and control or reference responses, (2) the variance among treatment and control replicate experimental units, (3) the number of replicate units for the treatment and control samples, (4) the number of animals exposed within a replicate exposure chamber, and (5) the selected probabilities of Type I (α) and Type II (β) errors.

12.2.3.4 Sample size or number of replicates may be fixed due to cost or space considerations, or may be varied to achieve a priori probabilities of α and β. The MDD should be established ahead of time based upon biological and program considerations. The investigator has little control of the variance among replicate exposure chambers. However, this variance component can be minimized by selecting test organisms that are as biologically similar as possible and maintaining test conditions within prescribed quality control (QC) limits.

12.2.3.5 The MDD is expressed as a percentage change from the mean control response. To test the equality of the control and treatment responses, a two-sample t-test with its associated assumptions is the appropriate parametric analysis. If the desired MDD, the number of replicates per treatment, the number of organisms per replicate and an estimate of typical among replicate variability, such as the coefficient of variation (CV) from a control sample, are available, it is possible to use a graphical approach as in Figure 12.1 to determine how likely it is that a 20% reduction will be detected in the treatment response relative to the control response. The CV is defined as 100% × (standard deviation divided by the mean). In a test design with eight replicates per treatment and with an α level of 0.05, high power (i.e., >0.8) to detect a 20% reduction from the control mean occurs only if the CV is 15% or less (Figure 12.1). The choice of these variables also affects the power of the test. If 5 replicates are used per treatment (Figure 12.2), the CV needs to be 10% or lower to detect a 20% reduction in response relative to the control mean with a power of 90%.

12.2.3.6 Relaxing the α level of a statistical test increases the power of the test. Figure 12.3 duplicates figure 12.1 except that α is 0.10 instead of 0.05. Selection of the appropriate α level of a test is a function of the costs associated with making Type I and II statistical errors. Evaluation of Figure 12.1 illustrates that with a CV of 15% and an α level of 0.05, there is an 80% probability (power) of detecting a 20% reduction in the mean treatment response relative to the control mean. However, if α is set at 0.10 and the CV remains at 15%, then there is a 90% probability (power) of detecting a 20% reduction relative to the control mean. The latter example would be preferable if an environmentally conservative analysis and interpretation of the data is desirable.

12.2.3.7 Increasing the number of replicates per treatment will increase the power to detect a 20% reduction in treatment response relative to the control mean. Note, however, that for less than eight replicates per treatment it is difficult to have high power (i.e., >0.80) unless the CV is less than 15%. If space or cost limit the number of replicates to fewer than eight per treatment, then it may be necessary to find ways to reduce the among replicate variability and consequently the CV. Options that are available include selecting more uniform organisms to reduce biological variability or increasing the α level of the test. For CVs in the range of 30% to 40%, even eight replicates per treatment is inadequate to detect small reductions ($\leq$20%) in response relative to the control mean.

DECISION	TR = Control	TR > Control
TR = Control	Correct $1 - \alpha$	Type II Error β
TR > Control	Type I Error α	Correct $1 - \beta$ (Power)

Figure 12.1 Treatment response (TR), alpha (α) represents the probability of making a Type I statistical error (false positive), beta (β) represents the probability of making a Type II statistical error (false negative).

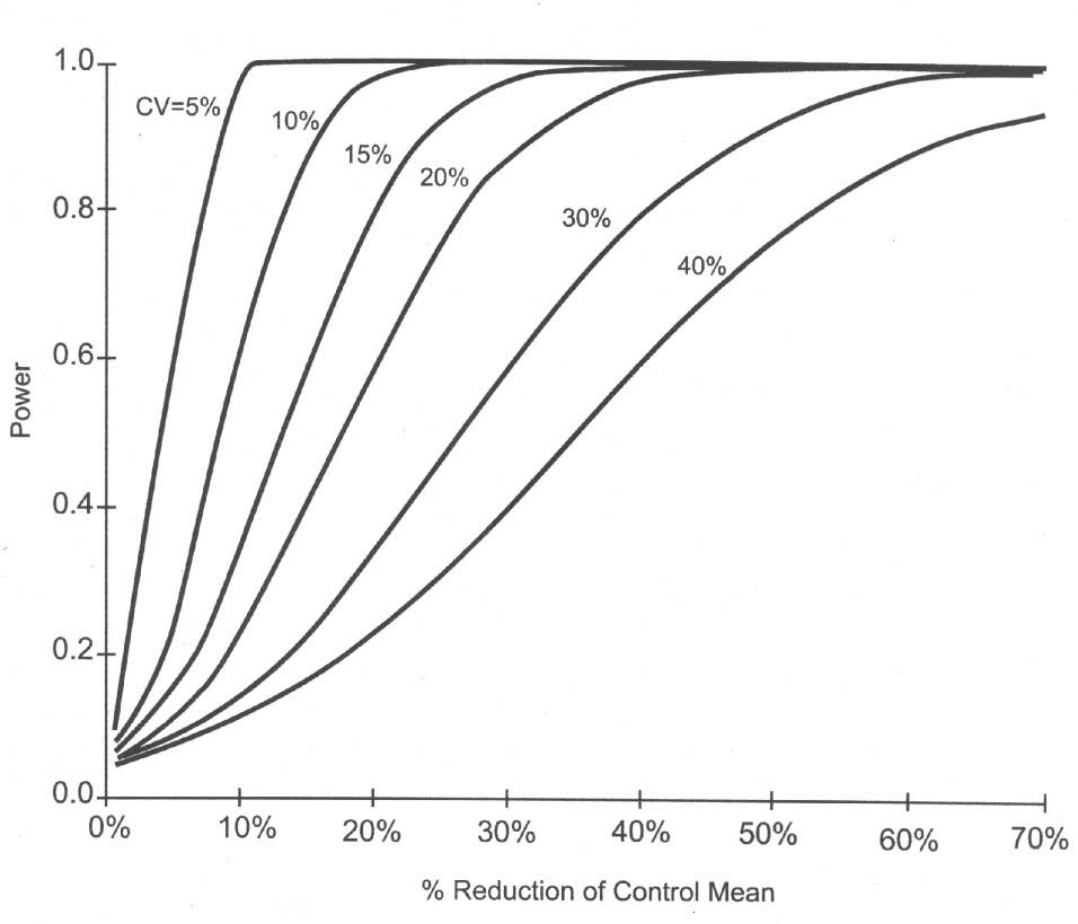

Figure 12.2 Power of the test versus percent reduction of the control mean at various CVs (eight replicates, alpha = 0.05 [one-tailed]).

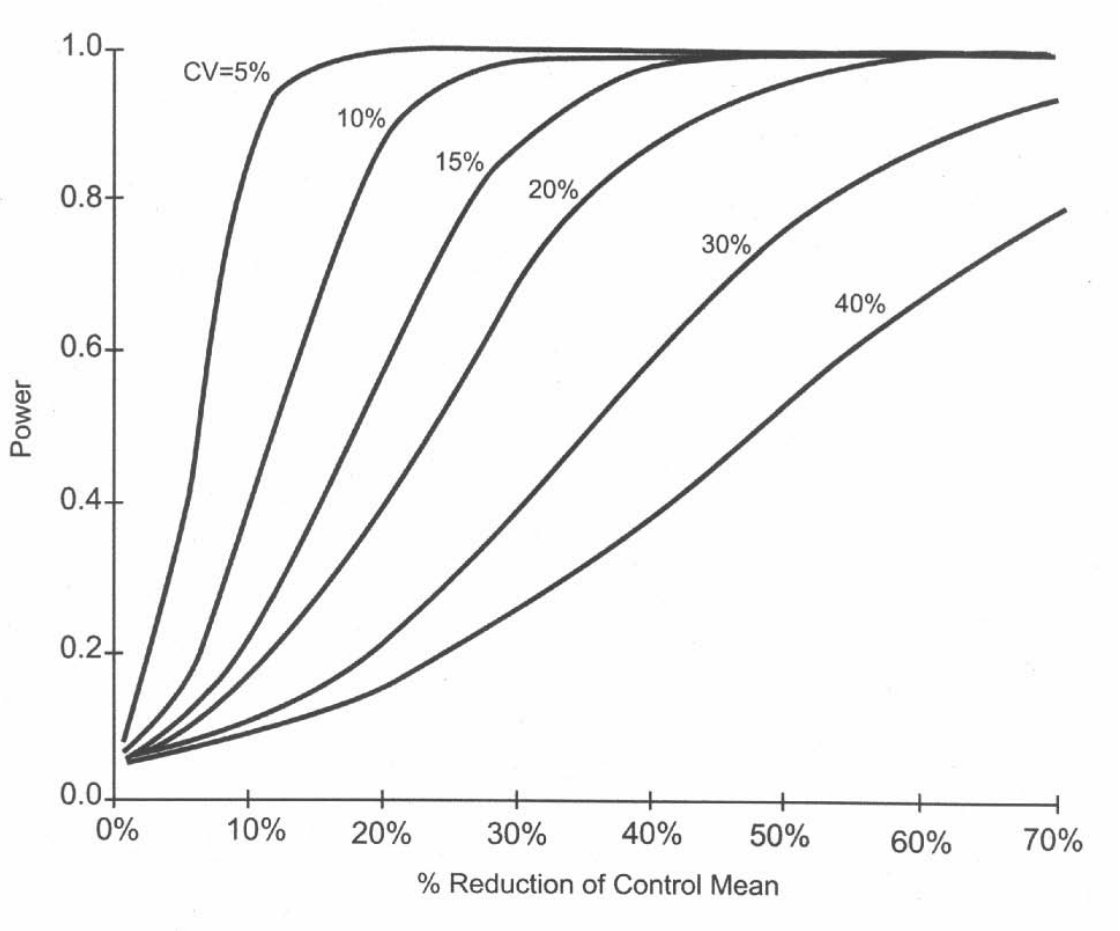

Figure 12.3 Power of the test versus percent reduction of the control mean at various CVs (five replicates, alpha = 0.05 [one-tailed]).

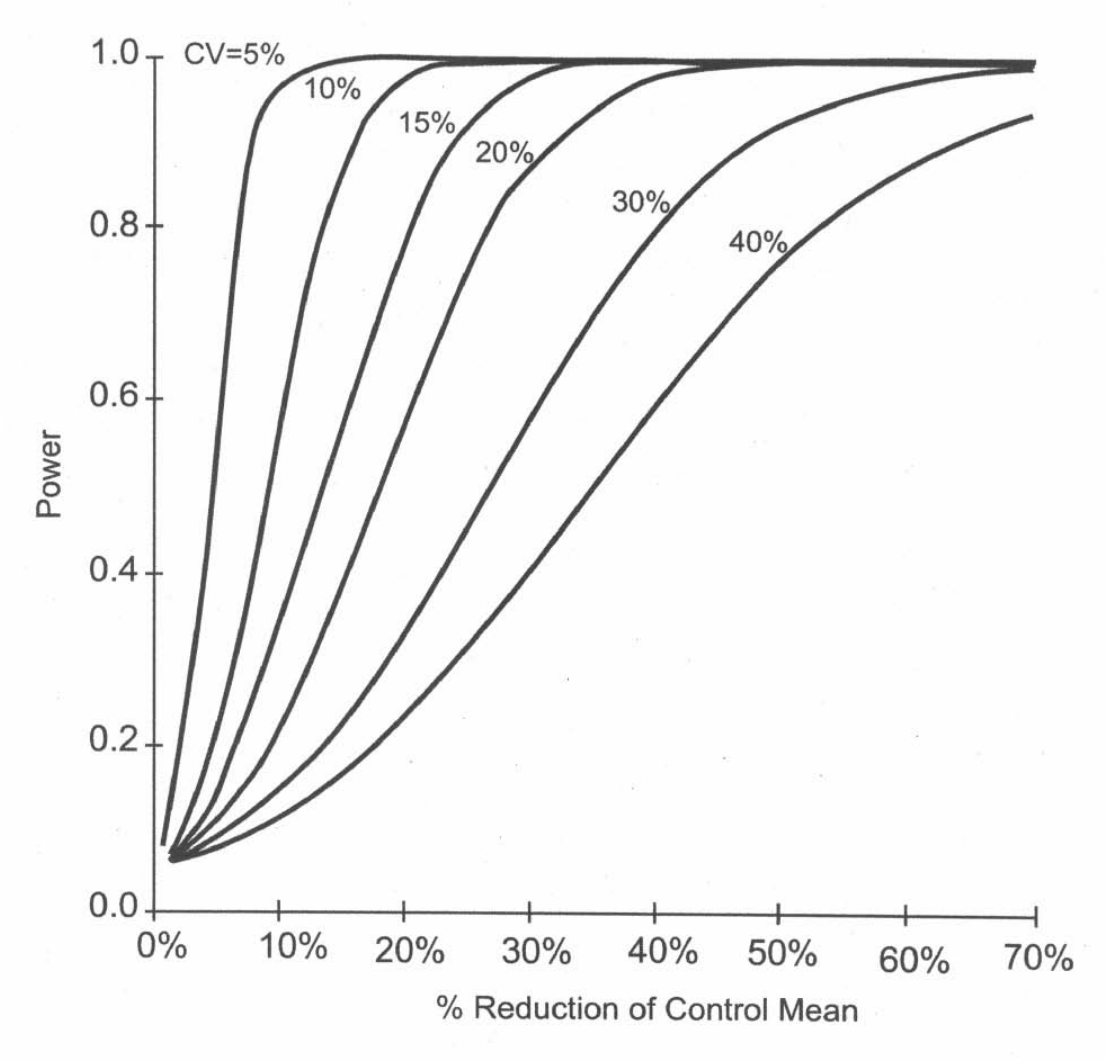

Figure 12.4 Power of the test versus percent reduction of the control mean at various CVs (eight replicates, alpha = 0.10 [one-tailed]).

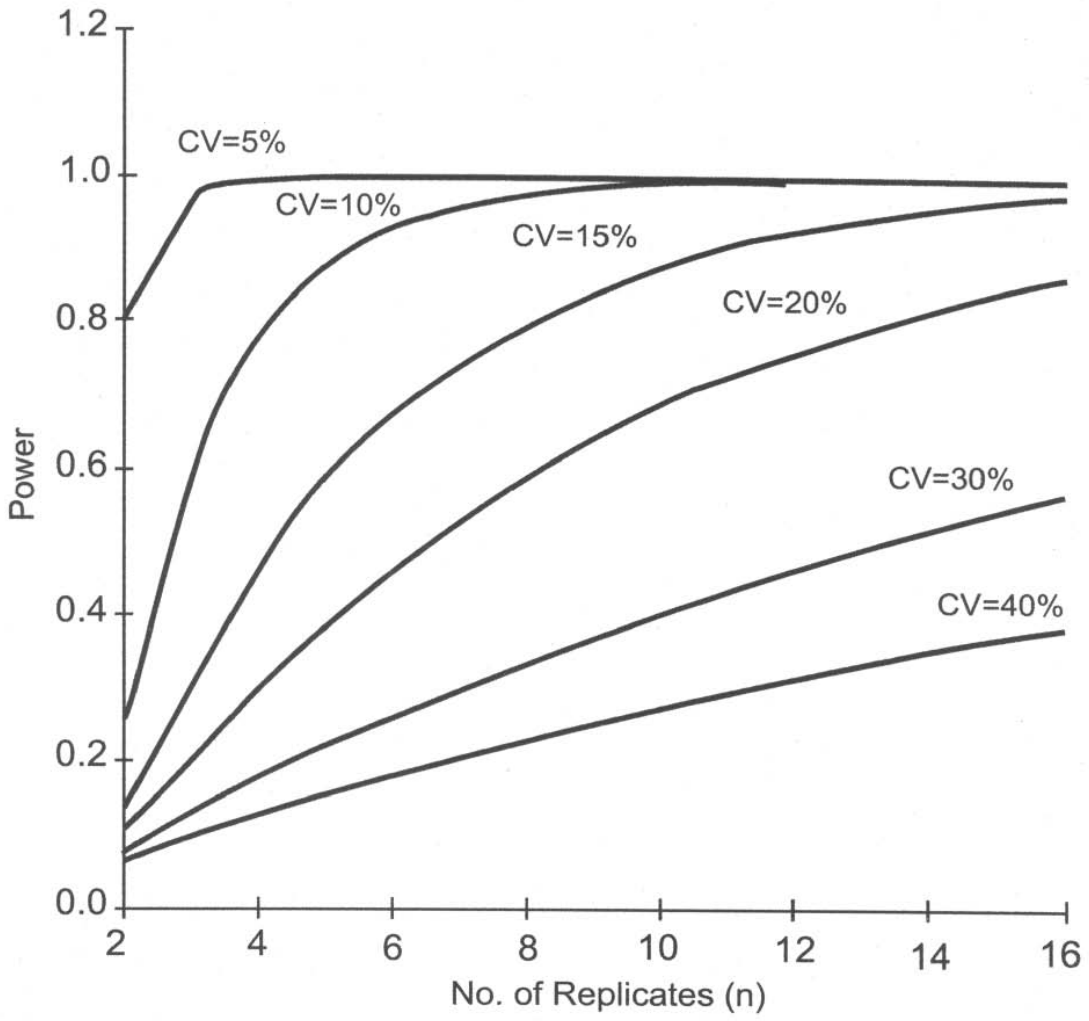

Figure 12.5 Effect of CV and number of replicates on the power to detect a 20% decrease relative to the control mean (alpha = 0.05 [one-tailed]).

12.2.3.8 The effect of the choice of α and β on number of replicates for various CVs is illustrated in Figure 12.6, in which the combined total probability of Type I and Type II statistical errors is fixed and assumed to be 0.25. An α of 0.10, therefore, establishes a β of 0.15. In Figure 12.6, if α = β = 0.125, the number of replicates required to detect a difference of 20% relative to the control is at a minimum. As α or β decrease, the number of replicates required to detect the same 20% difference relative to the control increases. However, the curves are relatively flat over the range of 0.05 to 0.20, and the curves are very dependent upon the choice of the combined total of α + β. Limiting the total of α + β to 0.10 greatly increases the number of replicates necessary to detect a preselected percentage reduction in mean treatment response relative to the control mean.

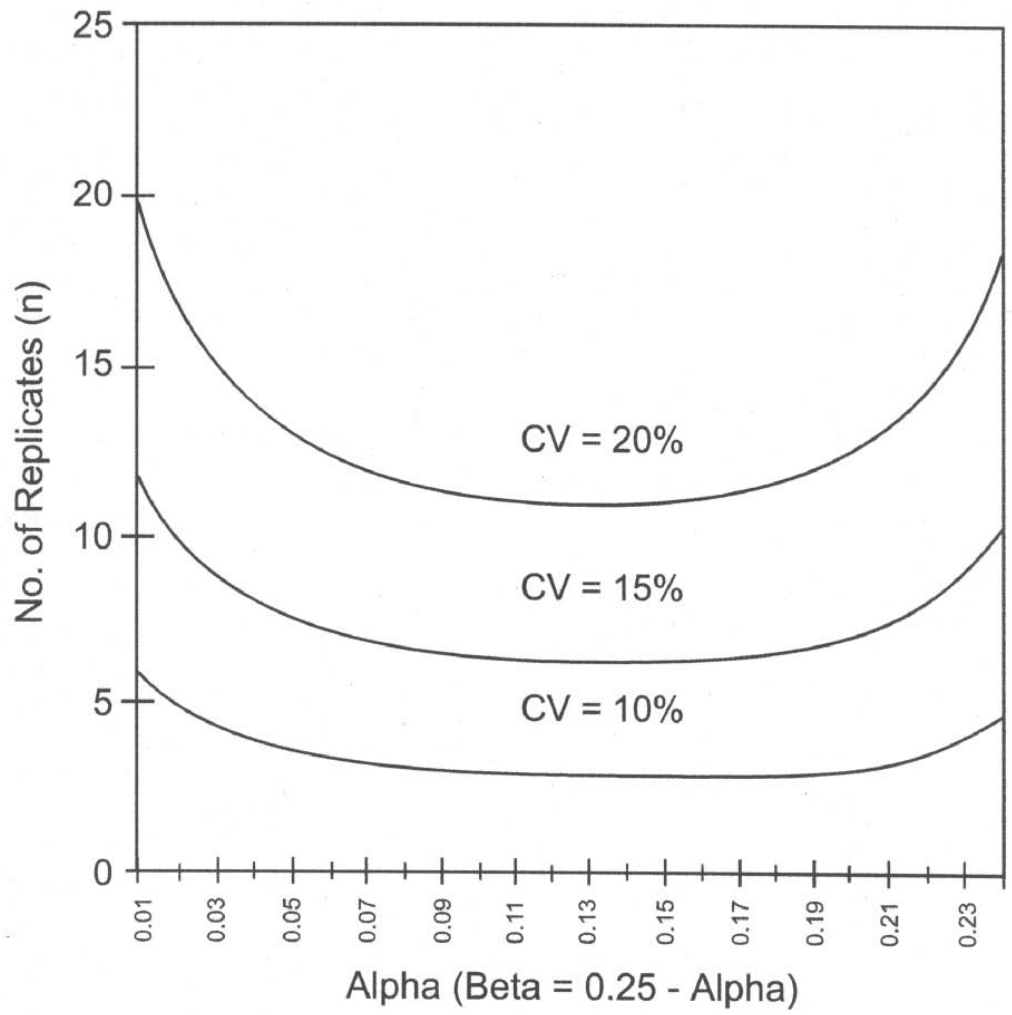

Figure 12.6 Effect of alpha and beta on the number of replicates at various CVs (assuming combined alpha + beta = 0.25).

12.2.4 Figure 12.7 outlines a decision tree for analysis of survival and growth data. In the tests described herein, samples or observations refer to replicates of treatments. Sample size n is the number of replicates (i.e., exposure chambers) in an individual treatment, not the number of organisms in an exposure chamber. Overall sample size "N" is the combined total number of replicates in all treatments. The statistical methods discussed in this section are described in general statistics texts such as Steel and Torrie (1980), Sokal and Rohlf (1981), Dixon and Massey (1983), Zar (1984), and Snedecor and Cochran (1989). It is recommended that users of this manual have at least one of these texts and associated statistical tables on hand. A nonparametric statistics text such as Conover (1980) may also be helpful.

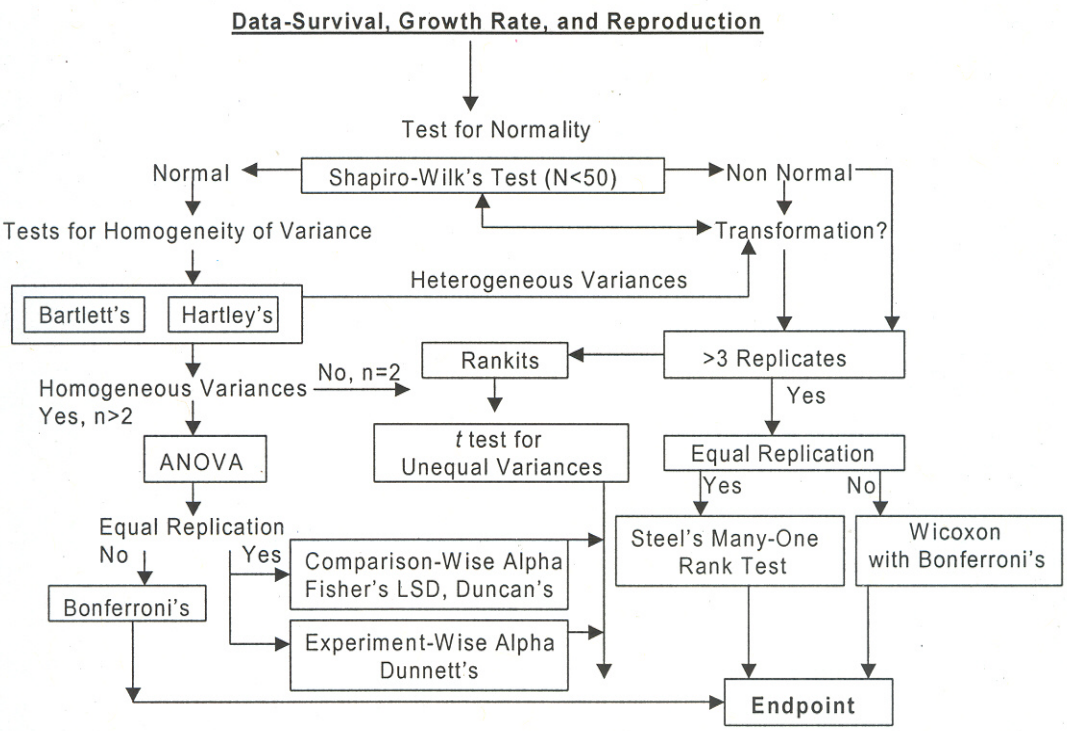

Figure 12.7 Decision tree for analysis survival and growth data.

12.2.4.1 Mean. The sample mean (x) is the average value, or $\Sigma x_i/n$, where:

n = number of observations (replicates)
x_i = ith observation
Σx_i = every × summed = $x_1 + x_2 + x_3 + \ldots + x_n$.

12.2.4.2 Standard Deviation. The sample standard deviation (s) is a measure of the variation of the data around the mean and is equivalent to $\sqrt{s^2}$. The sample variance, s^2, is given by the following "machine" or "calculation" formula:

$$s^2 = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n - 1}$$

12.2.4.3 Standard Error of the Mean. The standard error of the mean (SE, or $s/\sqrt{n}$) estimates variation among sample means rather than among individual values. The SE is an estimate of the SD among means that would be obtained from several samples of n observations each. Most of the statistical tests in this manual compare means with other means (e.g., dredged sediment mean with reference mean) or with a fixed standard (e.g., FDA action level; Lee et al. 1994). Therefore, the "natural" or "random" variation of sample means (estimated by SE), rather than the variation among individual observations (estimated by s), is required for the tests.

12.2.4.4 Tests of Assumptions. In general, parametric statistical analyses such as t-tests and analysis of variance are appropriate only if (1) there are independent, replicate experimental units for each treatment, (2) the observations within each treatment follow a normal distribution, and (3) variances for both treatments are equal or similar. The first assumption is an essential component of experimental design. The second and third assumptions can be tested using the data obtained from the experiment. Therefore, before conducting statistical analyses, tests for normality and equality of variances should be performed.

12.2.4.4.1 Outliers (extreme values) and systematic departures from a normal distribution (e.g., a log-normal distribution) are the most common causes of departures from normality or equality of variances. An outlier is an inconsistent or questionable data point that appears unrepresentative of the general trend exhibited by the majority of the data. Outliers may be detected by tabulation of the data, plotting, or by analysis of residuals. An explanation should be sought for any questionable data points. Without an explanation, data points should only be discarded with extreme caution. If there is no explanation, the analysis should be performed both with and without the outlier, and the results of both analyses should be reported. An appropriate transformation, such as the arcsine square root transformation, will normalize many distributions (USEPA 1985). Problems with outliers can usually be solved only by using nonparametric tests, but careful laboratory practices can reduce the frequency of outliers.

12.2.4.4.2 Tests for Normality. The most commonly used test for normality for small sample sizes (N<50) is the Shapiro-Wilk's Test. This test determines if residuals are normally distributed. Residuals are the differences between individual observations and the treatment mean. Residuals, rather than raw observations, are tested because subtracting the treatment mean removes any differences among treatments. This scales the observations so that the mean of residuals for each treatment and over all treatments is zero. The Shapiro-Wilk's Test provides a test statistic W, which is compared to values of W expected from a normal distribution. W will generally vary between 0.3 and 1.0, with lower values indicating greater departure from normality. Because normality is desired, one looks for a high value of W with an associated probability greater than the prespecified α level.

12.2.4.4.3 Table 12.2 provides α levels to determine whether departures from normality are significant. Normality should be rejected when the probability associated with W (or other normality test statistic) is less than α for the appropriate total number of replicates (N) and design. A balanced design means that all treatments have an equal number (n) of replicate exposure chambers. A design is considered unbalanced when the treatment with the largest number of replicates (n_{max}) has at least twice as many replicates as the treatment with the fewest replicates (n_{min}). Note that higher α levels are used when the number of replicates is small, or when the design is unbalanced, because these are the cases in which departures from normality have the greatest effects on t-tests and other parametric comparisons. If data fail the test for normality, even after transformation, nonparametric tests should be used for additional analyses.

12.2.4.4.4 Tables of quantiles of W can be found in Shapiro and Wilk (1965), Gill (1978), Conover (1980), USEPA (1989b), and other statistical texts. These references also provide methods of calculating W, although the calculations can be tedious. For that reason, commonly available computer programs or statistical packages are preferred for the calculation of W.

12.2.4.4.5 Tests for Homogeneity of Variances. There are a number of tests for equality of variances. Some of these tests are sensitive to departures from normality, which is why a test for normality should be performed first. Bartlett's Test or other tests such as Levene's Test or Cochran's Test (Winer 1971, Snedecor and Cochran 1989) all have similar power for small,

equal sample sizes (n=5) (Conover et al. 1980), and any one of these tests is adequate for the analyses in this section. Many software packages for t-tests and analysis of variance (ANOVA) provide at least one of the tests. Bartlett's Test is recommended for routine evaluation of homogeneity of variances (USEPA 1985, 1993b,c).

12.2.4.4.6 If no tests for equality of variances are included in the available statistical software, Hartley's F_{max} can easily be calculated:

$$F_{max} = (\text{ larger of } s_1^2, s_2^2) \ / \ (\text{ smaller of } s_1^2, s_2^2).$$

When F_{max} is large, the hypothesis of equal variances is more likely to be rejected. F_{max} is a two-tailed test because it does not matter which variance is expected to be larger. Some statistical texts provide critical values of F_{max} (Winer 1971, Gill 1978, Rohlf and Sokal 1981).

12.2.4.4.7 Levels of α for tests of equality of variances are provided in Table 12.2. These levels depend upon number of replicates in a treatment (n) and allotment of replicates among treatments. Relatively high α's (i.e., $\geq$0.10) are recommended because the power of the above tests for equality of variances is rather low (about 0.3) when n is small. Equality of variances is rejected if the probability associated with the test statistic is less than the appropriate α.

12.2.4.5 Transformations of the Data. When the assumptions of normality or homogeneity of variance are not met, transformations of the data may remedy the problem, so that the data can be analyzed by parametric procedures, rather than by a nonparametric technique. The first step in these analyses is to transform the responses, expressed as the proportion surviving, by the arcsine square root transformation. The arcsine-square root transformation is commonly used on proportionality data to stabilize the variance and satisfy the normality requirement. If the data do not meet the assumption of normality and there are four or more replicates per group, then the nonparametric test, Wilcoxon Rank Sum Test, can be used to analyze the data. If the data meet the assumption of normality, Bartlett's Test or Hartley's F test for equality of variances is used to test the homogeneity of variance assumption. Failure of the homogeneity of variance assumption leads to the use of a modified t test and the degrees of freedom for the test are adjusted.

12.2.4.5.1 The arcsine-square root transformation consists of determining the angle (in radians) represented by a sine value. In this transformation, the proportion surviving is taken as the sine value, the root of the sine value is calculated, and the angle (in radians) for the square root of the sine value is determined. When the proportion surviving is 0 or 1, a special modification of the transformation should be used (Bartlett 1937). An example of the arcsine-square root transformation and modification are provided below.

1. Calculate the response proportion (RP) for each replicate within a group,

   where RP = (number of surviving organisms)/(number exposed) .

2. Transform each RP to arcsine, as follows.

   a. For RPs greater than zero or less than one:

$$Angle\,(in\ radians) = arc\ sine\sqrt{(RP)}$$

   b. Modification of the arcsine when RP = 0:

$$Angle\,(in\ radians) = arc\ sine\sqrt{\frac{1}{4n}}$$

where n = number animals/treatment rep.

c. Modification of the arcsine when RP = 1.0:

Angle = 1.5708 radians – (radians for RP = 0)

12.2.4.6 Two Sample Comparisons (N=2). The true population mean (μ) and standard deviation (σ) are known only after sampling the entire population. In most cases samples are taken randomly from the population, and the s calculated from those samples is only an estimate of σ. Student's t-values account for this uncertainty. The degrees of freedom for the test, which are defined as the sample size minus one (n-1), should be used to obtain the correct t-value. Student t-values decrease with increasing sample size because larger samples provide a more precise estimate of μ and σ.

12.2.4.6.1 When using a t table, it is crucial to determine whether the table is based on one-tailed probabilities or two-tailed probabilities. In formulating a statistical hypothesis, the alternative hypothesis can be one-sided (one-tailed test) or two-sided (two-tailed test). The null hypothesis (Ho) is always that the two values being analyzed are equal. A one-sided alternative hypothesis (Ha) is that there is a specified relationship between the two values (e.g., one value is greater than the other) versus a two-sided alternative hypothesis (Ha), which is that the two values are simply different (i.e., either larger or smaller). A one-tailed test is used when there is an a priori reason to test for a specific relationship between two means such as the alternative hypothesis that the treatment mortality or tissue residue is greater than the control mortality or tissue residue. In contrast, the two-tailed test is used when the direction of the difference is not important or cannot be assumed before testing.

12.2.4.6.2 Since control organism mortality or tissue residues and sediment contaminant concentrations are presumed lower than reference or treatment sediment values, conducting one-tailed tests is recommended in most cases. For the same number of replicates, one-tailed tests are more likely to detect statistically significant differences between treatments (e.g., have a greater power). This is a critical consideration when dealing with a small number of replicates (such as eight per treatment). The other alternative for increasing statistical power is to increase the number of replicates, which increases the cost of the test.

12.2.4.6.3 There are cases when a one-tailed test is inappropriate. When no a priori assumption can be made as to how the values vary in relationship to one another, a two-tailed test should be used. An example of an alternative two-sided hypothesis is that the reference sediment total organic carbon (TOC) content is different (greater or lesser) from the control sediment TOC.

12.2.4.6.4 The t-value for a one-tailed probability may be found in a two-tailed table by looking up t under the column for twice the desired one-tailed probability. For example, the one-tailed t-value for α = 0.05 and df = 20 is 1.725 and is found in a two-tailed table using the column for α = 0.10.

12.2.4.7 The usual statistical test for comparing two independent samples is the two-sample

t-test (Snedecor and Cochran 1989). The t-statistic for testing the equality of means 0_1 and 0_2 from two independent samples with n_1 and n_2 replicates and unequal variances is:

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{s_1^2/n_1 + s_2^2/n_2} .$$

where s_1^2 and s_2^2 are the sample variances of the two groups. Although the equation assumes that the variances of the two groups are unequal, it is equally useful for situations in which the variances of the two groups are equal. This statistic is compared with the Student t distribution with degrees of freedom (df) given by Satterthwaite's (1946) approximation:

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2 / (n_1 - 1) + (s_2^2/n_2)^2 / (n_2 - 1)}$$

This formula can result in fractional degrees of freedom, in which case one should round the degree of freedom down to the nearest integer in order to use a t table. Using this approach, the degrees of freedom for this test will be less than the degrees of freedom for a t-test assuming equal variances. If there are unequal numbers of replicates in the treatments, the t-test with Bonferroni's adjustment can be used for data analysis (USEPA 1993b,c). When variances are equal, an F test for equality is unnecessary.

12.2.4.8 Nonparametric Tests. Tests such as the t-test, which analyze the original or transformed data, and which rely on the properties of the normal distribution, are referred to as parametric tests. Nonparametric tests, which do not require normally distributed data, analyze the ranks of data and generally compare medians rather than means. The median of a sample is the middle or 50th percentile observation when the data are ranked from smallest to largest. In many cases, nonparametric tests can be performed simply by converting the data to ranks or normalized ranks (rankits) and conducting the usual parametric test procedures on the ranks or rankits.

12.2.4.8.1 Nonparametric tests are useful because of their generality but have less statistical power than corresponding parametric tests when the parametric test assumptions are met. If parametric tests are not appropriate for comparisons because the normality assumption is not met, data should be converted to normalized ranks (rankits). Rankits are simply the z-scores expected for the rank in a normal distribution. Thus, using rankits imposes a normal distribution over all the data, although not necessarily within each treatment. Rankits can be obtained by ranking the data, then converting the ranks to rankits using the following formula:

$$rankit = z_{[(rank - 0.375) / (N + 0.25)]}$$

where z is the normal deviate and N is the total number of observations. Alternatively, rankits may be obtained from standard statistical tables such as Rohlf and Sokal (1981).

12.2.4.8.2 If normalized ranks are calculated, the ranks should be converted to rankits using the formula above. In comparisons involving only two treatments (N = 2), there is no need to test assumptions on the rankits or ranks; simply proceed with a one-tailed t-test for unequal variances using the rankits or ranks.

12.2.4.9 Analysis of Variance (N > 2). Some experiments are set up to compare more than one

treatment with a control while others may also be interested in comparing the treatments with one another. The basic design of these experiments is the same as for experiments evaluating pairwise comparisons. After the applicable comparisons are determined, the data must be tested for normality to determine if parametric statistics are appropriate and whether the variances of the treatments are equal. If normality of the data and equal variances are established, then an analysis of variance (ANOVA) may be performed to address the hypothesis that all the treatments including the control are equal. If normality or equality of variance are not established then transformations of the data may be appropriate or nonparametric statistics can be used to test for equal means. Tests for normality of the data should be performed on the treatment residuals. A residual is defined as the observed value minus the treatment mean, that is, $r_{ik} = o_{ik}$ - (k^{th} treatment mean). Pooling residuals provides an adequate sample size to test the data for normality.

12.2.4.9.1 The variances of the treatments should also be tested for equality. Currently there is no easy way to test for equality of the treatment means using analysis of variance if the variances are not equal. In a toxicity test with several treatments, one treatment may have 100% mortality in all of its replicates, or the control treatment may have 100% survival in all of its replicates. These responses result in 0 variance for a treatment, which results in a rejection of equality of variance in these cases. No transformation will change this outcome. In this case, the replicate responses for the treatment with 0 variance should be removed before testing for equality of variances. Only those treatments that do not have 0 replicate variance should be used in the ANOVA to get an estimate of the within treatment variance. After a variance estimate is obtained, the means of the treatments with 0 variance may be tested against the other treatment means using the appropriate mean comparison. Equality of variances among the treatments can be evaluated with the Hartley F_{max} test or Bartlett's test. The option of using nonparametric statistics on the entire set of data is also an alternative.

12.2.4.9.2 If the data are not normally distributed or the variances among treatments are not homogeneous, even after data transformation, nonparametric analyses are appropriate. If there are four or more replicates per treatment and the number of replicates per treatment is equal, the data can be analyzed with Steel's Many-One Rank test. Unequal replication among treatments requires data analysis with the Wilcoxon Rank Sum test with Bonferroni's adjustment. Steel's Many-One Rank test is a nonparametric test for comparing treatments with a control. This test is an alternative to the Dunnett's Procedure, and may be applied to data when the normality assumption has not been met. Steel's test requires equal variances across treatments and the control, but is thought to be fairly insensitive to deviations from this condition (USEPA 1993a). Wilcoxon's Rank Sum Tests is a nonparametric test to be used as an alternative to the Steel's test when the number of replicates are not the same within each treatment. A Bonferroni's adjustment of the pairwise error rate for comparison of each treatment versus the control is used to set an upper bound of alpha on the overall error rate. This is in contrast to the Steel's test with a fixed overall error rate for alpha. Thus, Steel's tests is a more powerful test (USEPA 1993a).

12.2.4.9.3 Different mean comparison tests are used depending on whether an α percent comparison-wise error rate or an α percent experiment-wise error rate is desired. The choice of a comparison-wise or experiment-wise error rate depends on whether a decision is based on a pairwise comparison (comparison-wise) or from a set of comparisons (experiment-wise). For example, a comparison-wise error rate would be used for deciding which stations along a gradient were acceptable or not acceptable, relative to a control or reference sediment. Each individual comparison is performed independently at a smaller α (than used in an experiment-wise comparison) such that the probability of making a Type I error in the entire series of comparisons

is not greater than the chosen experiment-wise α level of the test. This results in a more conservative test when comparing any particular sample to the control or reference. However, if several samples were taken from the same area and the decision to accept or reject the area was based upon all comparisons with a reference then an experiment-wise error rate should be used. When an experiment-wise error rate is used, the power to detect real differences between any two means decreases as a function of the number of treatment means being compared to the control treatment.

12.2.4.9.4 The recommended procedure for pairwise comparisons that have a comparison-wise α error rate and equal replication is to do an ANOVA followed by a one-sided Fisher's Least Significant Difference (LSD) test (Steel and Torrie 1980). A Duncan's mean comparison test should give results similar to the LSD. If the treatments do not contain equal numbers of replicates, the appropriate analysis is the t-test with Bonferroni's adjustment. For comparisons that maintain an experiment-wide α error rate, Dunnett's test is recommended for comparisons with the control.

12.2.4.9.5 Dunnett's test has an overall error rate of α, which accounts for the multiple comparisons with the control. Dunnett's procedure uses a pooled estimate of the variance, which is equal to the error value calculated in an ANOVA. Dunnett's procedure can only be used when the same number of replicate test chambers have been used at each treatment and the control.

12.2.4.9.6 To perform the individual comparisons, calculate the t statistic for each treatment and control combination, as follows:

$$t_i = \frac{(\overline{Y_1} - \overline{Y_i})}{S_w\sqrt{(1/n_1) + (1/n_i)}}$$

where: Y_i = Mean for each treatment
Y_1 = Mean for the control
S_w = Square root of the within mean square
n_1 = Number of replicates in the control
n_i = Number of replicates for treatment "i".

To quantify the sensitivity of the Dunnett's test, the minimum significant difference (MSD = MDD) may be calculated with the following formula:

$$MSD = d\,S_w\sqrt{(1/n_1) + (1/n)}$$

where: d = Critical value for the Dunnett's Procedure
S_w = The square root of the within mean square
n = The number of replicates per treatment, assuming an equal number of replicates at all treatment concentrations
n_1 = Number of replicates in the control.

12.2.5 Methods for Calculating LC50, EC50, and ICp.

12.2.5.1 Figure 12.8 outlines a decision for analysis of point estimate data. USEPA (USEPA 1985, USEPA 1989b, USEPA 1993b,c) discuss in detail the mechanics of calculating LC50 (or

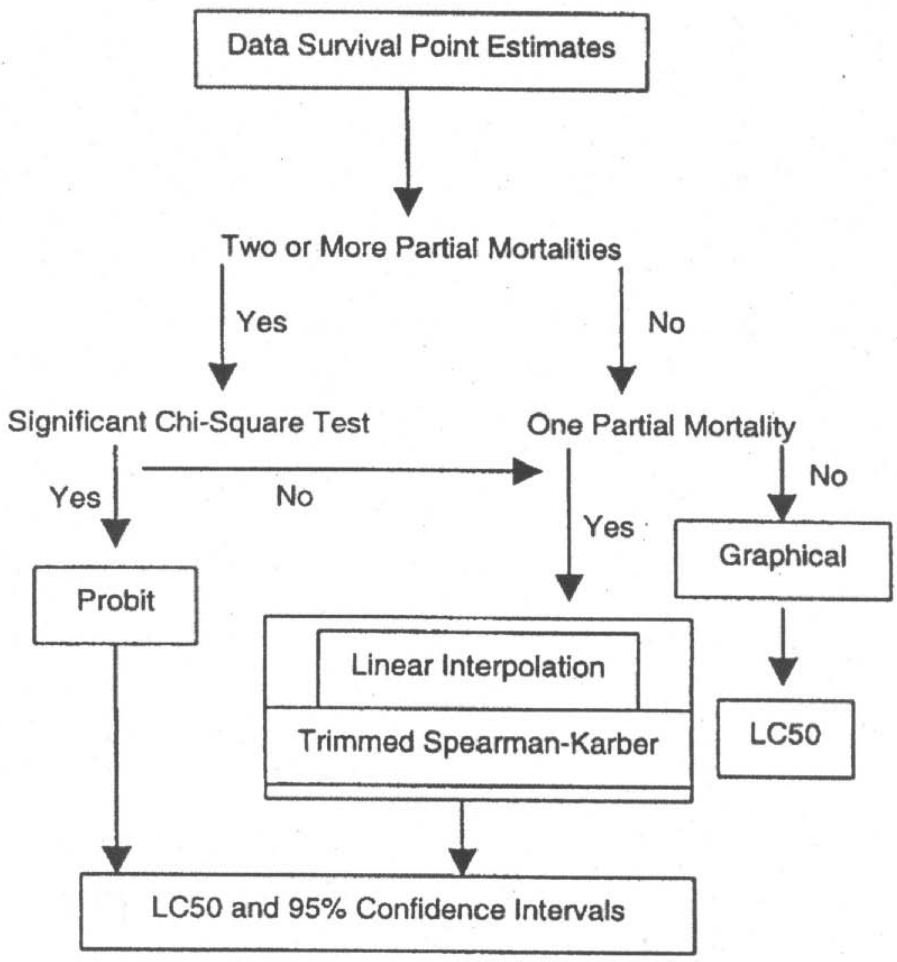

Figure 12.8 Decision tree for analysis of point estimate data.

EC50) or ICp values using the most current methods. The most commonly used methods are the Graphical, Probit, Trimmed Spearman-Karber and the Linear Interpolation Methods. In general, results from these methods should yield similar estimates. Each method is outlined below and recommendations presented for the use of each method.

12.2.5.2 Data for at least five test concentrations and the control should be available to calculate an LC50 although each method can be used with fewer concentrations. Survival in the lowest concentration must be at least 50% and an LC50 should not be calculated unless at least 50% of the organisms die in at least one of the serial dilutions. When less than 50% mortality occurs in the highest test concentration, the LC50 is expressed as greater than the highest test concentration.

12.2.5.3 Due to the intensive nature of the calculations for the estimated LC50 and associated 95% confidence interval using most of the following methods, it is recommended that the data be analyzed with the aid of computer software. A computer program to estimate the LC50 values and associated 95% confidence intervals with the methods discussed below

(except for the Graphical Method) was developed by USEPA and can be obtained at http://www.epa.gov/nerleerd/stat2.htm#tsk.

12.2.5.4 Graphical Method. This procedure estimates an LC50 (or EC50) by linearly interpolating between points of a plot of observed percentage mortality versus the base 10 logarithm ($\log_{10}$) of treatment concentration. The only requirement for its use is that treatment mortalities bracket 50%.

12.2.5.4.1 For an analysis using the Graphical Method the data should first be smoothed and adjusted for mortality in the control replicates. The procedure for smoothing and adjusting the data is detailed in the following steps: Let $p_0, p_1, ..., p_k$ denote the observed proportion mortalities for the control and the k treatments. The first step is to smooth the p_i if they do not satisfy $p_0 \leq p_1 \leq ... \leq p_k$. The smoothing process replaces any adjacent p_i's that do not conform to $p_0 \leq p_1 \leq ... \leq p_k$ with their average. For example, if p_i is less than p_{i-1} then:

$$P_{-1}^{s} = p_1^{s} = (p_i + p_{i-1}) / 2,$$

where p_1^{s} = the smoothed observed proportion mortality for concentration i.

Adjust the smoothed observed proportion mortality in each treatment for mortality in the control group using Abbott's formula (Finney 1971). The adjustment takes the form:

$$p_1^{a} = (p_1^{s} - p_0^{s}) / (1 - p_0^{s})$$

where: p_0^{s} = the smoothed observed proportion mortality for the control
p_1^{s} = the smoothed observed proportion mortality for concentration i.

12.2.5.5 Probit. The Probit Method is a parametric statistical procedure for estimating the LC50 (or EC50) and the associated 95% confidence interval (Finney 1971). The analysis consists of transforming the observed proportion mortalities with a probit transformation, and transforming the treatment concentrations to $\log_{10}$. Given the assumption of normality for the $\log_{10}$ of the tolerances, the relationship between the transformed variables mentioned above is about linear. This relationship allows estimation of linear regression parameters, using an iterative approach. A probit is the same as a z-score: for example, the probit corresponding to 70% mortality is $z_{.70}$ or = .52. The LC50 is calculated from the regression and is the concentration associated with 50% mortality or z = 0. To obtain a reasonably precise estimate of the LC50 with the Probit Method, the observed proportion mortalities must bracket 0.5 and the $\log_{10}$ of the tolerance should be normally distributed. To calculate the LC50 estimate and associated 95% confidence interval, two or more of the observed proportion mortalities must be between zero and one. The original percentage mortalities should be corrected for control mortality using Abbott's formula before the probit transformation is applied to the data.

12.2.5.5.1 A goodness-of-fit procedure with the Chi-square statistic is used to determine if the data fit the probit model. If many data sets are to be compared to one another, the probit method is not recommended because it may not be appropriate for many of the data sets. This method also is only appropriate for mortality data sets and should not be used for estimating end points that are a function of the control response, such as inhibition of growth. Most computer programs that generate probit estimates also generate confidence interval estimates

for the LC50. These confidence interval estimates on the LC50 may not be correct if replicate mortalities are pooled to obtain a mean treatment response. This can be avoided by entering the probit-transformed replicate responses and doing a least squares regression on the transformed data.

12.2.5.6 Trimmed Spearman-Karber. The Trimmed Spearman-Karber Method is a modification of the Spearman-Karber, nonparametric statistical procedure for estimating the LC50 and the associated 95% confidence interval (Hamilton et al. 1977). This procedure estimates the trimmed mean of the distribution of the $\log_{10}$ of the tolerance. If the log tolerance distribution is symmetric, this estimate of the trimmed mean is equivalent to an estimate of the median of the log tolerance distribution. Use of the Trimmed Spearman-Karber Method is only appropriate when the requirements for the Probit Method are not met (USEPA 1993b,c). This method is only appropriate for lethality data sets.

12.2.5.6.1 To calculate the LC50 estimate with the Trimmed Spearman-Karber Method, the smoothed, adjusted, observed proportion mortalities must bracket 0.5. To calculate a confidence interval for the LC50 estimate, one or more of the smoothed, adjusted, observed proportion mortalities must be between zero and one.

12.2.5.6.2 Smooth the observed proportion mortalities as described for the Probit Method. Adjust the smoothed observed proportion mortality in each concentration for mortality in the control group using Abbott's formula (see Probit Method). Calculate the amount of trim to use in the estimation of the LC50 as follows:

$$\text{Trim} = \max(p_1^a, 1 - p_k^a),$$

where: p_1^a = the smoothed, adjusted proportion mortality for the lowest treatment concentration, exclusive of the control

p_k^a = the smoothed, adjusted proportion mortality for the highest treatment concentration

k = the number of treatment concentrations, exclusive of the control.

12.2.5.7 Linear Interpolation Method. This method calculates a toxicant concentration that causes a given percent reduction (e.g., 25%, 50%, etc.) in the end point of interest and is reported as an ICp value (IC = inhibition concentration; where p = the percent effect). The procedure was designed for general applicability in the analysis of data from chronic toxicity tests, and the generation of an end point from a continuous model that allows a traditional quantitative assessment of the precision of the end point, such as confidence limits for the end point of a single test, and a mean and coefficient of variation for the end points of multiple tests.

12.2.5.7.1 As described in USEPA (USEPA 1993b,c), the Linear Interpolation Method of calculating an ICp assumes that the responses (1) are monotonically nonincreasing, where the mean response for each higher concentration is less than or equal to the mean response for the previous concentration, (2) follow a piecewise linear response function, and (3) are from a random, independent, and representative sample of test data. If the data are not monotonically nonincreasing, they are adjusted by smoothing (averaging). In cases where the responses at the low toxicant concentrations are much higher than in the controls, the smoothing process may result in a large upward adjustment in the control mean. In the Linear Interpolation Method,

the smoothed response means are used to obtain the ICp estimate reported for the test. No assumption is made about the distribution of the data except that the data within a group being resampled are independent and identically distributed.

12.2.5.7.2 The Linear Interpolation Method assumes a linear response from one concentration to the next. Thus, the IC is estimated by linear interpolation between two concentrations whose responses bracket the response of interest, the (p) percent reduction from the control.

12.2.5.7.3 If the assumption of monotonicity of test results is met, the observed response means (Y_i) should stay the same or decrease as the toxicant concentration increases. If the means do not decrease monotonically, the responses are "smoothed" by averaging (pooling) adjacent means. Observed means at each concentration are considered in order of increasing concentration, starting with the control mean (Y_1). If the mean observed response at the lowest toxicant concentration (Y_2) is equal to or smaller than the control mean (Y_1), it is used as the response. If it is larger than the control mean, it is averaged with the control, and this average is used for both the control response (M_1) and the lowest toxicant concentration response (M_2). This mean is then compared to the mean observed response for the next higher toxicant concentration (Y_3). Again, if the mean observed response for the next higher toxicant concentration is smaller than the mean of the control and the lowest toxicant concentration, it is used as the response. If it is higher than the mean of the first two, it is averaged with the first two, and the mean is used as the response for the control and two lowest concentrations of toxicant. This process is continued for data from the remaining toxicant concentrations. Unusual patterns in the deviations from monotonicity may require an additional step of smoothing. Where Y_i decrease monotonically, the Y_i become M_i without smoothing.

12.2.5.7.4 To obtain the ICp estimate, determine the concentrations C_J and C_{J+1} that bracket the response M_1 (1 - p/100), where M_1 is the smoothed control mean response, and p is the percent reduction in response relative to the control response. These calculations can easily be done by hand or with a computer program as described below. The linear interpolation estimate is calculated as follows:

$$ICp = C_J + [\, M_1\,(1 - p/100) - M_J \,] \; \frac{(C_{J+1} - C_J)}{(M_{J+1} - M_J)}$$

Where: C_J = tested concentration whose observed mean response is greater than $M_1(1 - p/100)$

C_{J+1} = tested concentration whose observed mean response is less than $M_1(1 - p/100)$

M_1 = smoothed mean response for the control

M_J = smoothed mean response for concentration J

M_{J+1} = smoothed mean response for concentration J + 1

P = percent reduction in response relative to the control response

IC_p = estimated concentration at which there is a percent reduction from the smoothed mean control response.

12.2.5.7.5 Standard statistical methods for calculating confidence intervals are not applicable for the ICp. The bootstrap method, as proposed by Efron (1982), is used to obtain the 95% confidence interval for the true mean. In the bootstrap method, the test data Y_{ji} is randomly resampled with replacement to produce a new set of data Y_{ji}*, that is statistically equivalent to the original data, but which produces a new and slightly different estimate of the ICp (ICp*). This process is repeated at least 80 times (Marcus and Holtzman 1988) resulting in multiple "data" sets, each with an associated ICp* estimate. The distribution of the ICp* estimates derived from the sets of resampled data approximates the sampling distribution of the ICp estimate. The standard error of the ICp is estimated by the standard deviation of the individual ICp* estimates. Empirical confidence intervals are derived from the quantiles of the ICp* empirical distribution. For example, if the test data are resampled a minimum of 80 times, the empirical 2.5% and the 97.5% confidence limits are about the second smallest and second largest ICp* estimates (Marcus and Holtzman 1988). The width of the confidence intervals calculated by the bootstrap method is related to the variability of the data. When confidence intervals are wide, the reliability of the IC estimate is in question. However, narrow intervals do not necessarily indicate that the estimate is highly reliable because of undetected violations of assumptions and the fact that the confidence limits based on the empirical quantiles of a bootstrap distribution of 80 samples may be unstable.

12.3 Data Interpretation

12.3.1 Sediments spiked with known concentrations of contaminants can be used to establish cause and effect relationships between chemicals and biological responses. Results of toxicity tests with test materials spiked into sediments at different concentrations can be reported in terms of an LC50 (median lethal concentration), an EC50 (median effect concentration), an IC50 (inhibition concentration), or as a NOEC (no observed effect concentration) or LOEC (lowest observed effect concentration) (section 3). Consistent spiking procedures should be followed in order to make interlaboratory comparisons (section 8.3).

12.3.2 Evaluating effect concentrations for chemicals in sediment requires knowledge of the factors controlling bioavailability. Similar concentrations of a chemical in units of mass of chemical per mass of sediment dry weight often exhibit a range of toxicity in different sediments (Di Toro et al. 1991, USEPA 1992c). Effect concentrations of chemicals in sediments have been correlated to interstitial water concentrations, and effect concentrations in water are often similar to effect concentrations in water-only exposures. The bioavailability of non-ionic organic compounds are often inversely related with the organic carbon concentration of the sediment. Whatever the route of exposure, the correlations of effect concentrations to interstitial water concentrations indicate predicted or measured concentrations in the interstitial water can be useful for quantifying the exposure concentration to an organism. Therefore, information on partitioning of chemicals between solid and liquid phases of sediment may be useful in establishing effect concentrations.

12.3.3 Toxic units can be used to help interpret the response of organisms to multiple contaminants in the sediment. A toxic unit is the concentration of a chemical divided by an effect concentration. For example, a toxic unit of exposure can be calculated by dividing the measured concentration of a chemical in pore water by the water-only LC50 for the same chemical (Ankley et al. 1994). Toxicity expressed as toxic units may be summed and this may provide information on the toxicity of complex mixtures (Ankley et al. 1994).

12.3.4 Field surveys can be designed to provide either a qualitative reconnaissance of the distribution of sediment contamination or a quantitative statistical comparison of contamination among sites (Burton and Ingersoll, 1994). Surveys of sediment toxicity are usually part of more comprehensive analyses of biological, chemical, geological, and hydrographic data. Statistical correlation can be improved and costs reduced if subsamples are taken simultaneously for sediment toxicity or bioaccumulation tests, chemical analyses, and benthic community structure.

12.3.5 Descriptive methods such as toxicity tests with field collected sediment should not be used alone to evaluate sediment contamination. An integration of several methods using the weight of evidence is needed to assess the effects of contaminants associated with sediment. Hazard evaluations integrating data from laboratory exposures, chemical analyses, and benthic community assessments provide strong complementary evidence of the degree of pollution-induced degradation in aquatic communities (Chapman et al. 1992, Burton 1991).

12.3.6 Toxicity Identification Evaluation (TIE) procedures can be used to help provide insights as to the specific contaminants responsible for toxicity in sediment (USEPA, 1991a; Ankley and Thomas, 1992). For example, the toxicity of contaminants such as metals, ammonia, hydrogen sulfide, and non-ionic organic compounds can be identified using TIE procedures.

12.4 Data Reporting

12.4.1 The record of results of an acceptable sediment test should include the following information either directly or by referencing available documents.

12.4.1.1 Name of test and investigator(s), name and location of the laboratory, and the dates of the start and end of the test.

12.4.1.2 Source of control or test sediment, method for collection, handling, shipping, storage, and disposal of sediment.

12.4.1.3 Source of test material, lot number if applicable, composition (identities and concentrations of major ingredients and impurities if known), known chemical and physical properties, and the identity and concentration(s) of any solvent used.

12.4.1.4 Source and characteristics of overlying water, description of any pretreatment, and results of any demonstration of the ability of an organism to survive or grow in the water.

12.4.1.5 Source, history, and age of test organisms; source, history and age of brood stock, culture procedures; and source and date of collection of the test organisms, scientific name, name of the person who identified the organisms and the taxonomic key used, age or life stage, means and ranges of weight or length, observed diseases or unusual appearance, treatments, and holding procedures.

12.4.1.6 Source and composition of food; concentrations of test material and other contaminants; procedure used to prepare food; and feeding methods, frequency, and ration.

12.4.1.7 Description of the experimental design and test chambers, the depth and volume of sediment and overlying water in the test chambers, lighting, number of test chambers and the number of test organisms per treatment, date and time test starts and ends, temperature measurements, dissolved oxygen concentration (as percent saturation), and any aeration used before starting the test and during the conduct of the test.

12.4.1.8 Methods used for physical and chemical characterization of sediment.

12.4.1.9 Definition(s) of the effects used to calculate LC50 or EC50s and biological end points for tests.

12.4.1.10 A table of the biological data for each test chamber for each treatment including the control(s), in sufficient detail to allow independent statistical analysis.

12.4.1.11 Methods used for statistical analysis of the data.

12.4.1.12 Summary of general observations or other effects or symptoms.

12.4.1.13 Anything unusual about the test, any deviation from these procedures, and any other relevant information.

12.4.2 Published reports should contain enough information to clearly identify the methodology used and the quality of the results.

Section 13

Precision and Accuracy

13.1 Determining Precision and Accuracy

13.1.1 Precision is a term that describes the degree to which data generated from replicate measurements differ and reflects the closeness of agreement between randomly selected test results. Accuracy is the difference between the value of the measured data and the true value and is the closeness of agreement between an observed value and an accepted reference value. Quantitative determination of precision and accuracy in sediment testing of aquatic organisms is difficult or may be impossible in some cases, as compared to analytical (chemical) determinations. This is due, in part, to the many unknown variables that affect organism response. Determining the accuracy of a sediment test using field samples is not possible since the true values are not known. Since there is no acceptable reference material suitable for determining the accuracy of sediment tests, accuracy of the test methods has not been determined (section 13.2).

13.1.2 Sediment tests exhibit variability due to several factors (section 9). Test variability can be described in terms of two types of precision: single-laboratory (intralaboratory or repeatability; section 13.5.1) precision or multilaboratory (interlaboratory or reproducibility; section 13.5.2) precision (often measured with round-robin or ring tests). Intralaboratory precision reflects the ability of trained laboratory personnel to obtain consistent results repeatedly when performing the same test on the same organism using the same toxicant. Interlaboratory precision is a measure of how reproducible a method is when conducted by a large number of laboratories using the same method, organism, and toxic sample. Generally, intralaboratory results are less variable than interlaboratory results (USEPA 1991b, USEPA 1993a,b,c, Hall et al. 1989, Grothe and Kimerle 1985).

13.1.3 A measure of precision can be calculated using the mean and relative standard deviation (percent coefficient of variation, or CV% = standard deviation/mean × 100) of the calculated end points from the replicated end points of a test. However, precision reported as the CV should not be the only approach used for evaluating precision of tests and should not be used for the NOEC effect levels derived from statistical analyses of hypothesis testing. The CVs may be very high when testing extremely toxic or nontoxic samples. For example, if there are multiple replicates with no survival and one with low survival the CV may exceed 100%, yet the range of response is actually quite consistent. Therefore, additional estimates of precision should be used, such as range of responses, minimum detectable differences (MDD) compared to control survival or growth. Several factors can affect the precision of the test, including test organism age, condition, sensitivity, handling of the test organisms, overlying water quality, and the experience in conducting tests. For these reasons, it is recommended that trained laboratory personnel conduct the tests in accordance with the procedures outlined in section 9. Quality assurance practices should include (1) single laboratory precision determinations using reference toxicants for each of the test organisms that are used to determine the ability of the laboratory personnel to obtain precise results—these determinations should be made before conducting a sediment test and should be routinely performed as long as whole sediment tests are being conducted, (2) control charts (section 13.3) should be prepared for each reference toxicant and test organism to determine if the test results are within prescribed limits, and (3) tests must meet the minimum criteria of test acceptability specific for each test organism (Table 11.3; USEPA 1991b).

13.1.4 Intralaboratory precision data are routinely calculated for test organisms using water-only 96-h exposures to a reference toxicant, such as $CdCl_2$. Intralaboratory precision data should be tracked using a control chart. Each laboratory's reference toxicant data will reflect conditions unique to that facility, including dilution water, culturing, and other variables (section 9). However, each laboratory's reference toxicant CVs should reflect good repeatability.

13.1.5 No interlaboratory precision (round-robin) tests have been completed using 10-day whole sediment tests with *Mulinia lateralis.*

13.2 Accuracy

13.2.1 The accuracy of sediment toxicity tests cannot be determined since there is no acceptable reference material. The accuracy of the reference toxicity tests can only be evaluated by comparing test responses to control charts.

13.3 Replication and Test Sensitivity

13.3.1 The sensitivity of sediment tests will depend in part on the number of replicates per concentration, the probability levels (alpha and beta) selected, and the type of statistical analysis. For a given level of variability, the sensitivity of the test will increase as the number of replicates is increased. The minimum recommended number of replicates varies with the objectives of the test and the statistical method used for analysis of the data (section 12).

13.4 Demonstrating Acceptable Laboratory Performance

13.4.1 It is the responsibility of a laboratory to demonstrate its ability to obtain precise results with reference toxicants before it performs sediment tests (section 9.16). Intralaboratory precision, expressed as a coefficient of variation (CV), of the range for each type of test to be used in a laboratory should be determined by performing five or more tests with different batches of test organisms, using the same reference toxicant, at the same concentrations, with the same test conditions (e.g., the same test duration, type of water, age of test organisms, feeding), and same data analysis methods. This should be done to gain experience for the toxicity tests and a point of reference for future testing. A reference toxicant concentration series (0.5 or higher) should be selected that will consistently provide partial mortalities at two or more concentrations of the test chemical (section 9.14, Table 9.1).

13.4.2 The quality of test organisms obtained from an outside source, regardless of whether they are from culture or collected from the field, must be verified by conducting a reference-toxicity test concurrently with the sediment test. For cultured organisms, the supplier should provide data with the shipment describing the history of the sensitivity of organisms from the same source culture. For field-collected organisms, the supplier should provide data with the shipment describing the collection location, the water salinity and temperature at the time of collection, and collection site sediment for holding and acclimation purposes. If the supplier has not conducted five reference toxicity tests with the test organism, it is the responsibility of the testing laboratory to conduct these five reference toxicity tests before starting a sediment test (section 9.14.1).

13.4.3 Before conducting tests with contaminated sediment, the laboratory should demonstrate its ability to conduct tests by conducting five exposures in control sediment as outlined in Table

11.1. It is recommended that these five exposures with control sediment be conducted concurrently with the five reference toxicity tests described in section 9.14.1.

13.4.4 A control chart should be prepared for each combination of reference toxicant and test organism. End points from five tests are adequate for establishing the control charts. In this technique, a running plot is maintained for the values (X_i) from successive tests with a given reference toxicant , and the end points (LC50, NOEC, ICp) are examined to determine if they are within prescribed limits. Control charts as described in USEPA (2001) are used to evaluate the cumulative trend of results from a series of samples. The mean and upper and lower control limits (±2 SD) are recalculated with each successive test result. After 2 years of data collection, or a minimum of 20 data points, the control chart should be maintained using only the 20 most recent data points.

13.4.5 The outliers, which are values falling outside the upper and lower control limits, and trends of increasing or decreasing sensitivity, are readily identified using control charts. With an alpha of 0.05, one in 20 tests would be expected to fall outside of the control limits by chance alone. During a 30-day period, if 2 of 20 reference toxicity tests fall outside the control limits, the sediment toxicity tests conducted during the time in which the second reference toxicity test failed are suspect and should be considered as provisional and subject to careful review.

13.4.5.1 A sediment test may be acceptable if specified conditions of a reference toxicant test fall outside the expected ranges (section 9). Specifically, a sediment test should not automatically be judged unacceptable if the LC50 for a given reference toxicity test falls outside the expected range or if mortality in the control of the reference toxicity test exceeds 20%. All the performance criteria outlined in Table 11.3 must be considered when determining the acceptability of a sediment test. The acceptability of the sediment test would depend on the experience and judgment of the investigator and the regulatory authority.

13.4.6 If the value from a given test with the reference toxicant falls more than two standard deviations (SD) outside the expected range, the sensitivity of the organisms and the overall credibility of the test system may be suspect (USEPA 1993a). In this case, the test procedure should be examined for defects and could be repeated with a different batch of test organisms, if necessary.

13.4.7 Performance should improve with experience, and the control limits for point estimates should gradually narrow. However, control limits of ±2 SD, by definition, will be exceeded 5% of the time, regardless of how well a laboratory performs. Highly proficient laboratories that develop a very narrow control limit may be unfairly penalized if a test that falls just outside the control limits is rejected de facto. For this reason, the width of the control limits should be considered in determining whether or not an outlier is to be rejected. This determination may be made by the regulatory authority evaluating the data.

13.4.8 The recommended reference toxicity test consists of a control and five or more concentrations in which the end point is an estimate of the toxicant concentration that is lethal to 50% of the test organisms in the time period prescribed by the test. The LC50 is determined by an appropriate procedure, such as the Trimmed Spearman-Karber Method, Probit Method, Graphical Method, or the Linear Interpolation Method (section 12).

13.4.9 The point estimation analysis methods recommended in this manual have been chosen primarily because they are well tested, well documented, and are applicable to most types of test

data. Many other methods were considered in the selection process, and it is recognized that the methods selected are not the only possible methods of analysis for toxicity data.

13.5 Precision of Sediment Toxicity Test Methods

13.5.1 Intralaboratory Precision

13.5.1.1 Intralaboratory precision was evaluated for *Mulinia lateralis* by performing five reference toxicant tests with copper spiked sand. The mean LC50 was 31.7% copper spiked sand (estimated at 190 g/kg dry sand) with a standard deviation of 16.8 and a coefficient of variation of 53.0. The coefficient of variation is relatively high and is attributed to the occurrence of only one concentration tested that yielded partial mortalities in most tests. The range finder provided a range where partial mortalities were expected. However, subsequent tests did not show partial mortality at any but one of the concentrations chosen. The mean EC50 was 20.9% copper spiked sand (estimated at 125.4 g/Kg dry sand) with a standard deviation of 6.4 and a coefficient of variation of 30.6. Since the EC50 provides an estimate of an effect on each individual clam, the EC50 is not as affected by having only one concentration tested that yielded partial mortality and the coefficient of variation is smaller.

13.5.2 Interlaboratory Precision

13.5.2.1 Interlaboratory precision has not been evaluated for *Mulinia lateralis.*

References

Adams, W.U., R.A. Kimerle, and R.G. Mosher. 1985. Aquatic safety assessment of chemicals sorbed to sediments. In: *Aquatic Toxicology and Hazard Assessment: 7th Volume,* ASTM STP 854, R.D. Cardwell, R. Purdy, and R.C. Bahner, Eds., Philadelphia, PA: American Society for Testing and Materials, pp. 429-453.

ASTM. 1988. Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphipods. ASTM 1993 Annual Book of Standards Vol. 11.04, E729-88, Philadelphia, PA.

ASTM. 1992. Guide for conducting 10-day static sediment toxicity tests with marine and estuarine amphipods. ASTM 1992 Annual Book of Standards Vol. 13.04, E1367-92, Philadelphia, PA.

ASTM. 1993a. Standard terminology relating to biological effects and environmental fate. ASTM 1993 Annual Book of Standards Vol. 11.04, E943-93, Philadelphia, PA.

ASTM. 1993b. Standard guide for designing biological tests with sediments. ASTM 1993 Annual Book of Standards Vol. 11.04, E1525-93, Philadelphia, PA.

ASTM. 1994. Draft standard guide for the use of light in laboratory testing. Under development by B.M. Greenburg, Department of Biology, University of Waterloo, Waterloo, Ontario. ASTM Subcommittee E47.11, January, 1994.

ASTM. 2004. Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing. ASTM 2004 Annual Book of Standards Vol. 11.05, E1391-02, Philadelphia, PA.

ASTM. 2006. Standard guide for determination of the bioaccumulation of sediment-associated contaminants by benthic invertebrates. ASTM 2006 Annual Book of Standards Vol. 11.06, E1688-00a, Philadelphia, PA.

Ankley, G. and N. Thomas. 1992. Interstitial water toxicity identification evaluation approach. In: Sediment classification methods compendium, pp. 5-1 to 5-14. EPA 823-R-92-006, Washington, DC.

Ankley, G.T., Band, D.A., Balough, J.C., Reynoldson, T.B., Day, K.E., and Hoke, R.A. 1994. Evaluation of potential confounding factors in sediment toxicity tests with three freshwater benthic invertebrates. *Environmental Toxicology and Chemistry,* 13(4):627-635.

Barnard, J.L. and C.M. Barnard. 1982. The genus *Rhepoxynius* (Crustacea: Amphipoda: Phoxocephalidae) in American Seas. *Smithsonian Contributions to Zoology,* 357:1-49.

Bartlett, M.S. 1937. Some examples of statistical methods of research in agriculture and applied biology. *Journal, Royal Statistical Society, Supplement,* 4:137-183.

Berry, W.J., D.J. Hansen, J.D. Mahony, D.L. Robson, D.M. Di Toro, B.P. Shipley and B.A. Rogers. 1996. Predicting the Toxicity of Metal-Spiked Laboratory Sediments Using Acid-Volatile Sulfide and Interstitial Water Normalizations. Environmental Toxicology and Chemistry 15(12):2067-2079.

Boese B.L., H. Lee II, D.T. Specht, R.C. Randall, and M.H. Winsor. 1990. Comparison of aqueous and solid-phase uptake for hexachlorobenzene in the tellinid clam *Macoma nasuta* (Conrad): A mass balance approach. *Environmental Toxicology and Chemistry,* 9:221-231.

Bousfield, E.L. 1973. Shallow water gammaridean amphipoda of New England. Ithaca, NY. Cornell University Press, 312 pp.

Bureau of National Affairs, Inc. 1986. U.S. Environmental Protection Agency General Regulation for Hazardous Waste Management. Washington, DC.

Burgess, R.M. and G.E. Morrison. 1994. A short-exposure, sublethal, sediment toxicity test using the bivalve *Mulinia lateralis:* Statistic design and comparative sensitivity. *Environmental Toxicology and Chemistry.* 13(4), 571-580.

Burgess, R.M., B.A. Rogers, S.A. Rego, J.M. Corbin, and G.E. Morrison. 1994. Sand spiked with copper as a reference toxicant material for sediment toxicity testing: A preliminary evaluation. *Archives of Environmental Contamination and Toxicology,* 26:163-168.

Burton, G.A. 1991. Assessing the toxicity of freshwater sediments. *Environmental Toxicology and Chemistry,* 10:1585-1627.

Burton, G.A., Jr. and C.G. Ingersoll. 1994. Evaluation of sediment toxicity. In: The Assessment and Remediation of Contaminated Sediment (ACRS) Program. EPA 905-B94-002. USEPA, Region V, Chicago, IL, p. 86-121.

Burton, G.A., K.E. Day, and P.F. Landrum. 1994. Revision to ASTM E1391-90 Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing (08/05/93 draft in ASTM review).

Calabrese, A. 1969. The early life history and larval ecology of the coot clam, *Mulina lateralis* (Say) (Mactridae:Pelecypoda). Ph.D. thesis. University of Connecticut, Storrs, CT.

Calabrese, A. 1970. Reproductive cycle of the coot clam, *Mulinia lateralis* in long Island Sound. *The Veliger,* 12(3):265-126.

Calabrese A., E.W. Rhodes. 1974. Culture of *Mulinia lateralis* and *Crepidula fornicata* embryos and larvae for studies of pollution effects. Thalassia Jugoslavia, 10:89-102.

Carlson, A.R., G.L. Phipps, V.R. Mattson, P.A. Kosian, and A.M. Cotter. 1991. The role of acid-volatile sulfide in determining cadmium bioavailability and toxicity in freshwater sediments. *Environmental Toxicology and Chemistry,* 10:1309-1319.

Chapman, P.M. 1989. Current approaches to developing sediment quality criteria. *Environmental Toxicology and Chemistry,* 8:589-599.

Chapman, P.M., E.A. Power, and G.A. Burton, Jr. 1992. Integrative assessments in aquatic ecosystems. In: *Sediment toxicity assessment*. G.A. Burton, Jr., Ed. Ann Arbor, MI: Lewis Publishers, pp. 313-340.

Conover, W.J. 1980. Practical Nonparametric Statistics. 2nd Ed. John Wiley and Sons, New York, NY, 493 p.

Craig, G.R. 1984. Bioassessment of sediments: Review of previous methods and recommendations for future test protocols. IEC Beak consultants, Ltd. Mississauga, Ontario.

Cripe, G.M. 2006. Contaminated sediment testing with the bivalve, *Mulinia lateralis:* Culture refinement for organism availability. *Environmental Toxicology and Chemistry,* 25:1332-1336.

Davenport, R. and A. Spacie. 1991. Acute phototoxicity of harbor and tributary sediments from lower Lake Michigan. *Journal Great Lakes Research,* 17:51-56.

Day, K.E., R.S. Kirby, and T.B. Reynoldson. 1992. The effects of sediment manipulations on chronic sediment bioassays with three species of benthic invertebrates. Abstract. Presented at the 13'th Annual Meeting, Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 8-12.

DeWitt, T.H., G.R. Ditsworth, and R.C. Swartz. 1988. Effects of natural sediment features on survival of the Phoxocephalid amphipod, *Rhepoxynius abronius. Marine Environmental Research,* 25:99-124.

DeWitt, T.H., R.C. Swartz, and J.O. Lamberson. 1989. Measuring the acute toxicity of estuarine sediments. *Environmental Toxicology and Chemistry,* 8:1035-1048.

DeWitt, T.H., M.S. Redmond, J.E. Sewall, and R.C. Swartz. 1992. Development of a chronic sediment toxicity test for marine benthic amphipods. U.S. Environmental Protection Agency. CBP/TRS/89/93.

DeWoskin, R.S. 1984. Good laboratory practice regulations: A comparison. Research Triangle Institute, Research Triangle Park, NC, 63 p.

Dickson. K.L., A.W. Maki, and W.A. Brungs. 1984. Fate and effects of sediment-bound chemicals in aquatic systems. Pergamon Press, New York, NY.

Dillon, T.M. and A.B. Gibson. 1986. Bioassessment methodologies for the regulatory testing of freshwater dredged material. Miscellaneous Paper EL-86-6, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Di Toro, D.M., J.D. Mahony, D.J. Hansen, K.J. Scott, M.B. Hicks, S.M. Mays, and M.S. Redmond. 1990. Toxicity of cadmium in sediments: The role of acid volatile sulfide. *Environmental Toxicology and Chemistry* 9:1489-1504.

Di Toro, D.M., C.S Zarba, D.J. Hansen, W.J. Berry, R.C. Swartz, C.E. Cowan, S.P. Pavlou, H.E. Allen, N.A. Thomas, and P.R. Paquin. 1991. Technical basis for establishing sediment quality criteria for nonionic organic chemicals using equilibrium partitioning. *Environmental Toxicology and Chemistry,*10:1541-1583.

Ditsworth, G.R., D.W. Schults, and J.K.P. Jones. 1990. Preparation of benthic substrates for sediment toxicity testing. *Environmental Toxicology and Chemistry,* 9:1523-1529.

Dixon, W.J. and F.J. Massey, Jr. 1983. Introduction to statistical analysis. 4th Ed. McGraw-Hill Book Company, New York, NY, 678 p.

Efron, B. 1982. The Jackknife, the Bootstrap, and other resampling plans. CBMS 38, Soc. Industr. Appl. Math., Philadelphia, PA.

Environment Canada. 1992. Biological test method: Acute test for sediment toxicity using marine or estuarine amphipods. Environment Canada, Conservation and Protection, Ottowa, Ontario, 83 p., EPS 1/RM/26.

Fairweather, P.G. 1991. Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research,* 42:555-567.

Finney, D.J. 1971. *Probit analysis*. Third edition, Cambridge, University Press, London, 333 p., 1971.

Food and Drug Administration. 1978. Good laboratory practices for non-clinical laboratory studies. Part 58. Fed. Reg. 43(247):60013-60020 (December 22, 1978).

Frankenberg, D., S.L. Coles, and R.E. Johannes. 1967. The potential trophic significance of *Callianassa major* fecal pellets. *Limnology and Oceanography,* 12(1):113-120.

Gill, J.L. 1978. Design and analysis of experiments in the animal and medical sciences. Vol. 3. Appendices. The Iowa State University Press, Ames, IA, 173 p.

Green, R.H. 1979. *Sampling design and statistical methods for environmental biologists*. New York, NY: Wiley Interscience, 257 p.

Grothe, D.R. and R.A. Kimerle. 1985. Inter- and intra-laboratory variability in *Daphnia magna* effluent toxicity test results. *Environmental Toxicology and Chemistry,* 4:189-192.

Guillard, R.L. and J.H. Ryther. 1962. Studies of marine planktonic diatoms: *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology,* 8:229-239.

Hadley N.M, J.J. Manzi, A.G. Everson, R.T. Dillon, C.E. Battey. N.M. Peacock. 1997. A Manual for the Culture of the Hard Clam *Mercenaria* spp. in South Carolina. South Carolina Sea Grant Consortium, Charleston, SC.

Hall, W.S., J.B. Patoczka, R.J. Mirenda, B.A. Porter, and E. Miller. 1989. Acute toxicity of industrial surfactants to *Mysidopsis bahia. Archives of Environmental Contamination and Toxicology,* 18:765-772.

Hamilton, M.A., R.C. Russo, and R.V. Thurston. 1977. Trimmed Spearman-Karber method for estimating median lethal concentrations in toxicity bioassays. *Environmental Science and Technology,* 11:714-719. Correction 12:417 (1978).

Hanks, R.W. 1968. Benthic community formation in a "new" marine environment. *Chesapeake Science,* 9:163-172.

Hansen, D.J. 1984. Utility of toxicity tests to measure effects of substances on marine organisms. In: *Concepts in Marine Pollution*. Harris H. White, ed., pp 33-56.

Harkey, G.A., P.F. Landrum, and S.J. Klaine. 1994. Preliminary studies on the effect of feeding during whole sediment bioassays using *Chironomus riparius* larvae. *Chemosphere,* 28:597-606.

Holland, A.F., N.K. Mountford, M.H. Heigal, K.R. Kaumeyer, and J.A. Mihurski. 1980. Influence of predation on infaunal abundance in upper Chesapeake Bay, USA. *Marine Biology,* 57:221-235.

Hopkins, D.R. 1986. Atlas of the distributions and abundances of common benthic species in San Francisco Bay, California. water Resources Investigations Report 86-4003. Menlo Parl, CA: U.S. Geological Survey, 16 pp.

Huber, M., M.C. Pelletier, J.B. Charles and R.M. Burgess. 1997. Ammonia Tolerance of the Bivalve Mulinia lateralis Sublethal Sediment Toxicity Test. Environmental Contamination and Toxicology, 59:292-297.

Hurlbert, S.H. 1984. Psuedoreplication and the design of ecological field experiments. *Ecological Monographs,* 54(2):187-211.

Ingersoll, C.G. 1991. Sediment toxicity and bioaccumulation testing. *Standardization News* 19: 28-33.

Ingersoll, C.G. and M.K. Nelson. 1990. Testing sediment toxicity with *Hyalella azteca* (Amphipoda) and *Chironomus riparius* (Diptera). In: W.G. Landis and W.H. van der Schalie (eds.), *Aquatic toxicology and risk assessment*, 13th volume. ASTM STP 1096. Philadelphia, PA, pp. 93-109.

Ingersoll C.G., F.J. Dwyer, and T.W. May. 1990. Toxicity of inorganic and organic selenium to *Daphnia magna* (Cladocera) and *Chironomus riparius* (Diptera). *Environmental Toxicology and Chemistry* 9:1171-1181.

Ingersoll, C.G., D.R. Buckler, E.A. Crecelius, and T.W. La Point. 1993. U.S. Fish and Wildlife Service and Battelle final report for the USEPA and GLNPO assessment and remediation of contaminated sediment (ARCS) project: Biological assessment of contaminated Great Lakes sediment. EPA-905-R93 -006, Chicago, IL.

Karickhoff, S.W. and K.R. Morris. 1985. Sorption dynamics of hydrophobic pollutants in sediment suspensions. *Environmental Toxicology and Chemistry,* 4:469-479.

Kemble, N.E., W.G. Brumbaugh, E.L. Brunson, F.J. Dwyer, C.G. Ingersoll, C.G., D.P. Monda, and D.F. Woodward. 1994. Toxicity of metal-contaminated sediments from the upper Clark Fork River, MT, to aquatic invertebrates in laboratory exposures. *Environmental Toxicology and Chemistry,* in press.

Kemp, P.F. and R.C. Swartz. 1988. Acute toxicity of interstitial and particle-bound cadmium to a marine infaunal amphipod. *Marine Environmental Research,* 26:135-153.

Kemp, P.F., F.A. Code, and R.C. Swartz. 1985. Life history and production of the Phoxocephalid amphipod *Rhepoxynius abronius* (Barnard). *Journal of Crustacean Biology,* 5:449-464.

Kennedy, V.S. and J.A. Mihurski. 1972. Effects of temperature on the respiratory metabolism of three Chesapeake Bay bivalves. Chesapeake Science, 13(1):1-22.

Kennedy, V.S., W.H. Roosenburg, and M. Castagna. 1974. Temperature-time relationships for survival of embryos and larvae of *Mulinia lateralis* (Mollusca: Bivalvia). *Marine Biology,* 24:137-145.

Knezovich, J.P., F.L. Harrison, and R.G. Wilhelm. 1987. The bioavailability of sediment-sorbed organic chemicals: A review. *Water, Air, and Soil Pollution,* 32:233-245.

Lamberson, J.O. and R.C. Swartz. 1988. Use of bioassays in determining the toxicity of sediment to benthic organisms. In: *Toxic contaminants and ecosystem health: A Great Lakes focus*., M.S. Evans, Ed. New York, NY: John Wiley and Sons, pp. 257-279.

Lamberson, J.O., T.H. DeWitt, and R.C. Swartz. 1992. Assessment of sediment toxicity to marine benthos. In: *Sediment toxicity assessment*, G.A. Burton, Jr., Ed. Ann Arbor, MI: Lewis Publishers, pp. 183-211.

Landrum, P.F. 1989. Bioavailability and toxicokinetics of polycyclic aromatic hydrocarbons sorbed to sediments for the amphipod, *Pontoporeia hoyi. Environmental Science and Technology,* 23:588-595.

Landrum, P.F. and W.R. Faust. 1992. Variation in the bioavailability of polycyclic aromatic hydrocarbons sorbed to sediments for the amphipod *Pontoporeia hoyi. Environmental Toxicology and Chemistry,* 11:1197-1208.

Lee, D.R. 1980. Reference toxicants in quality control of aquatic bioassays. In: A.L. Buikema and J. Cairns Jr. (eds.), *Aquatic invertebrate bioassays*. ASTM STP 715, Philadelphia, PA, pp. 188-199.

Lee, H., II, B.L. Boese, J. Pelletier, M. Winsor, D.T. Specht, and R.C. Randall. 1989. Guidance manual: Bedded sediment bioaccumulation tests. USEPA Report No. 600/x-89/302, ERL-Narragansett, 232 p.

Long, E.R. and L.G. Morgan. 1990. The potential for biological effects of sediment-sorbed contaminants in the NOAA National Status and Trends program. NOAA Technical Memorandum NOS OMA 52, Seattle, WA, 175 pp + appendices.

Long, E.R., M.F. Buchman, S.M. Bay, R.J. Breteler, R.S. Carr, P.M. Chapman, J.E. Hose, A.L. Lissner, J. Scott, and D.A. Wolfe. 1990. Comparative evaluation of five sediment toxicity tests with sediments from San Francisco Bay and Tomales Bay, CA. *Environmental Toxicology and Chemistry,* 9:1193-1214.

Long, E.R., D.D. MacDonald, S.L. Smith, and F.D. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management,* 19(1):81-97.

Luchenbach, M.W. 1984. Effects of adult infauna on new recruits: implications for the role of biotic refuges. *Journal of Experimental Marine Biology and Ecology,* 105:197-206.

Marcus, A.H. and A.P. Holtzman. 1988. A robust statistical method for estimating effects concentrations in short-term fathead minnow toxicity tests. Manuscript submitted to the Criteria and Standards Division, U.S. Environmental Protection Agency, by Battelle Washington Environmental Program Office, Washington, DC, June 1988, under EPA Contract No. 69-03-3534. 39 p.

McGee, B.L., C.E. Schlekat, and E. Reinharz. 1993. Assessing sublethal levels of sediment contamination with the estuarine amphipod, *Leptocheirus plumulosus. Environmental Toxicology and Chemistry,* 12:577-588.

Morrison, G.E. and E. Petrocelli. 1991. Suitability of *Mulinia lateralis* as a euryhaline toxicity test species. *Canadian Technical Report of Fisheries and Aquatic Sciences,* 1774(1):337-340.

National Research Council (NRC). 1989. Contaminated marine sediments—Assessment and remediation, NRC, National Academy Press, Washington, DC.

Nichols, F.H. and J.K. Thompson. 1985. Persistence of and introduced mudflat community in south San Francisco Bay, California. *Marine Ecology Progress Series,* 24:83-97.

Oris, J.T. and J.P. Giesy. 1985. The photoenhanced toxicity of anthracene to juvenile sunfish *(Lepomis* spp.). *Aquatic Toxicology,* 6:133-146.

Petrocelli, E. and G.E. Morrison. 1990. Coot clam *(Mulinia lateralis)* embryo/larval toxicity test. ERL-N standard operation procedure 1.03.008

Pelletier, M.C., R.M. Burgess, K.T. Ho, A. Kuhn, R.A. McKinney and S.A. Ryba. 1997. Phototoxicity of Individual Polycyclic Aromatic Hydrocarbons and Petroleum To Marine Invertebrate Larvae and Juveniles. Environmental Toxicology and Chemistry, 16(10):2190-2199.

Pelletier, M.C., R.M. Burgess, M.G. Cantwell, J.R. Serbst, K.T. Ho and S.A. Ryba. 2000. Importance of Maternal Transfer of the Photoreactive Polycyclic Aromatic Hydrocarbon

Fluoranthene from Benthic Adult Bivalves to their Pelagic Larvae. Environmental Toxicology and Chemistry, 19(11):2691-2698.

Plumb, R.H. 1981. Procedures for handling and chemical analyses of sediment and water samples. U.S. Environmental Protection Agency/ Corps of Engineers Technical Committee on Criteria for Dredged and Fill Material. Contract EPA-4805572010, EPA/CE-81-1, 478 p.

Redmond, M.S. and K.J. Scott. 1989. Amphipod predation by the infaunal polychaete, *Nephtys incisa. Estuaries.* 12(3): 205-207.

Reish, D.J. 1988. The use of toxicity testing in marine environmental research, Chapter 10. In: D.F. Soule and G.S. Kleppel (eds.), *Marine organisms as indicators*. Springer-Verlag, New York, pp. 213-245.

Rhodes, E., A. Calabrese, and W. Cable. 1972. The development of methods for the rearing of the coot clam, *Mulinia lateralis,* and three species of coastal bivalves in the laboratory. In: *Culture of Marine Invertebrate Animals*. p. 273-277. W.L. Smith and M.H. Chandley, eds. Plenum Press.

Rohlf, F.J. and R.R. Sokal. 1981. *Statistical Tables*. W.H. Freeman and Company, New York, NY.

Sanders, H.L., P.C. Mangelsdorf, Jr., and G.R. Hampson. 1965. Salinity and faunal distribution in the Pocasset River, Massachusetts. *Limnology and Oceanography* Supplement to Vol. 10: R216 - R229.

Satterthwaite, F.W. 1946. An approximate distribution of estimates of variance components. *Biometrae Bulletin,* 2:110-114.

Schlekat, C.E., B.L. McGee, and E. Reinhartz. 1992. Testing sediment toxicity in Chesapeake Bay with the amphipod *Leptocheirus plumulosus:* an evaluation. *Environmental Toxicology and Chemistry,* 11:225-236.

Schults, D.W., S.P. Ferraro, L.M. Smith, F.A. Roberts, and C.K. Poindexter. 1992. A comparison of methods for collecting interstitial water for trace organic compounds and metal analyses. *Water Research,* 26:989-995.

Schuytema, G.S., A.V. Nebeker, W.L. Griffis, and C.E. Miller. 1989. Effects of freezing on sediments contaminated with DDT and endrin. *Environmental Toxicology and Chemistry,* 8:883-891.

Scott, K.J. 1989. Effects of contaminated sediments on marine benthic biota and communities. In: *Contaminated marine sediments—Assessment and remediation*. National Research Council. National Academy Press, Washington, DC. pp. 132-154.

Scott, K.J. and M.S. Redmond. 1989. The effects of a contaminated dredged material on laboratory populations of the tubicolous amphipod, *Ampelisca abdita.* In: *Aquatic Toxicology and Hazard Assessment: 12th Volume*, ASTM STP 1027, U.M. Cowgill and L.R. Williams, Eds. Philadelphia, PA: American Society for Testing and Materials. pp. 289-303.

Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika,* 52:591-611.

Sheehan, C.S. and M.C. Pelletier. 1992. Modification of the 48 hour embryo/larval test using the bivalve, *Mulinia lateralis.* Poster presentation to American Society of Testing and Materials. Pittsburg, PA.

Shuba, P.J., H.E. Tatem, and J.H. Carroll. 1978. Biological assessment methods to predict the impact of open-water disposal of dredged material. U.S. Army Corps of Engineers Technical Report D-78-5Q, Washington, DC.

Shumway, S.E. and R.C. Newell. 1984. Energy resource allocation in *Mulinia lateralis* (Say), an opportunistic bivalve from shallow water sediments. *Ophelia,* 23(2):101-118.

Smith, R.I. 1964. Keys to marine invertebrates of the Woods Hole region. Contribution No. 11. Systemics-Ecology Program, Marine Biological Laboratory. Woods Hole, MA.

Snedecor, G.W. and G.C. Cochran. 1989. *Statistical Methods*. 8th Ed. The Iowa State University Press, Ames, IA, 507 p.

Sokal, R.R. and F.J. Rohlf. 1981. *Biometry*, 2nd edition. W.H. Freeman and Company, New York.

Southerland, E., M. Kravitz, and T. Wall. 1992. Management framework for contaminated sediments (The USEPA Sediment Management Strategy). In: *Sediment Toxicity Assessment* G.A. Burton, Ed. Ann Arbor, MI: Lewis Publishers, pp. 341-370.

Steel, R.G.D. and J.H. Torrie. 1980. *Principles and procedures of statistics*. New York, NY: McGraw Hill Book Co.

Stein, J.R. 1973. *Handbook of phycological methods: Culture methods and growth measurements*. New York, NY: Cambridge University Press, 445 pp.

Stemmer, B.L., G.A. Burton Jr., and G. Sasson-Brickson. 1990a. Effect of sediment spatial variance and collection method on cladoceran toxicity and indigenous microbial activity determinations. *Environmental Toxicology and Chemistry,* 9:1035-1044.

Stemmer, B. L., Burton, Jr., G. A., and Leibfritz-Frederick, S. 1990b. Effect of sediment test variables on selenium toxicity to *Daphnia magna. Environmental Toxicology and Chemistry,* 9:1035-1044.

Swartz, R.C. 1989. Marine sediment toxicity tests. In: *Contaminated Marine Sediments—Assessments and Remediation*, Washington, D.C.: National Academy Press, pp. 115-129.

Swartz, R.C., W.A. DeBen, J.K.P. Jones, J.O. Lamberson, and F.A. Code. 1985. Phoxocephalid amphipod bioassay for marine sediment toxicity. In: Aquatic toxicology and hazard assessment: 7th Volume. ASTM STP 854, R.D. Cardwell, R. Purdy, and R.C. Bahner, Eds., Philadelphia, PA: American Society for Testing and Materials, pp. 284-307.

Swartz, R.C., P.F. Kemp, D.W. Schultz, and J.O. Lamberson. 1988. Effects of mixtures of sediment contaminants on the marine infaunal amphipod, *Rhepoxynius abronius. Environmental Toxicology and Chemistry,* 7:1013-1020.

Tagatz, M.E., G.R. Plaia, and C.H. Deans. 1985. Effects of 1,2,4-trichlorobenzene on estuarine macrobenthic communities exposed via water and sediment. *Ecotoxicology and Environmental Safety,* 10:35 1-360.

Taylor, J.K. 1987. *Quality assurance of chemical measurements*. Lewis Publ., Inc., Chelsea, MI.

Tomasovic, M.J., Dwyer, F.J., Greer, I.E., and Ingersoll, C.G. 1995. Recovery of known-age *Hyalella azteca* (Amphipoda) from sediment toxicity tests. *Environmental Toxicology and Chemistry,* 14(7):1177-1180.

USEPA. 1979a. Handbook for analytical quality assurance in water and wastewater laboratories. EPA-600/4-79-019, Cincinnati, OH.

USEPA. 1979b. Methods for chemical analysis of water and wastes. EPA-600/4-79-020, Cincinnati, OH.

USEPA. 1980a. Proposed good laboratory practice guidelines for toxicity testing. Paragraph 163.60-6. Fed. Reg. 45:26377-26382 (April 18, 1980).

USEPA. 1980b. Physical, chemical, persistence, and ecological effects testing; good laboratory practice standards (proposed rule). 40 CFR 772, Fed. Reg. 45:77353-77365 (November 21, 1980).

USEPA. 1985. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. 3rd Ed. EPA-600/4-85/013, Cincinnati, OH.

USEPA. 1986a. Quality criteria for water. EPA-440/5-86-001.

USEPA. 1986b. Occupational health and safety manual. Office of Administration, EPA, Washington, DC.

USEPA. 1988. Short-term methods for estimating chronic toxicity of effluents and receiving waters to marine and estuarine organisms. EPA/600/4-87/028. US EPA, Cincinnati, OH.

USEPA. 1989a. Evaluation of the apparent effects threshold (AET) approach for assessing sediment quality. Report of the sediment criteria subcommittee. SAB-EETFC-89-027, Washington, DC.

USEPA. 1989b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. EPA/600/4-89/001, Cincinnati, OH.

USEPA. 1990a. Evaluation of the equilibrium partitioning (EqP) approach for assessing sediment quality. Report of the sediment criteria subcommittee of the ecological processes and effects committee. EPA-SAB-EPEC-90-006, Washington, DC.

USEPA. 1990b. Evaluation of the sediment classification methods compendium. Report of the sediment criteria subcommittee of the ecological processes and effects committee. EPA-SABEPEC-90-01 8, Washington, DC.

USEPA. 1990c. Analytical procedures and quality assurance plan for determination of xenobiotic chemical contaminants in fish. EPA/600/3-90/023, Duluth, MN.

USEPA. 1990d. Analytical procedures and quality assurance plan for determination of PCDD/PCDF in fish. EPA/600/3-90/022, Duluth, MN.

USEPA. 1991a. Sediment toxicity identification evaluation: Phase I (Characterization), Phase II (Identification) and Phase III (Confirmation). Modifications of effluent procedures. EPA 600/6-91/007, Duluth, MN.

USEPA. 1991b. Technical support document for water quality-based toxic control. EPA 505/2-90/001, Washington, DC.

USEPA. 1992a. Proceedings from workshop on tiered testing issues for freshwater and marine sediments. Washington, DC, September 16-18.

USEPA. 1992b. An SAB report: Review of sediment criteria development methodology for non-ionic organic contaminants. Report of the sediment criteria subcommittee of the ecological processes and effects committee. EPA-SAB-EPEC-93-002, Washington, DC.

USEPA. 1992c. Sediment classification methods compendium. EPA 823-R-92-006, Washington, DC.

USEPA. 1993a. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. Fourth edition. EPA 600/4-90/027F, Cincinnati, OH.

USEPA. 1993b. Short-term methods for measuring the chronic toxicity of effluents and surface waters to freshwater organisms. Third edition. EPA 600/4-91/002, Cincinnati, OH.

USEPA. 1993c. Short-term methods for measuring the chronic toxicity of effluents and surface waters to marine and estuarine organisms. Second edition. EPA 600/4-91/002, Cincinnati, OH.

USEPA. 1994a. Procedures for conducting sediment toxicity and bioaccumulation tests using freshwater invertebrates. EPA-600/R-94/024. June 1994. USEPA, Office of Research and Development, Duluth, MN, 201 p.

USEPA. 1994b. Sediment Collection, Handling, Storage, and Manipulation. EPA-600/XXXX. USEPA, Office of Science and Technology, Washington, DC. In review.

USEPA. 2001. Method for Assessing the Chronic Toxicity of Marine and Estuarine Sediment-associated Contaminants with the Amphipod *Leptocheirus plumulosus*. First edition. EPA-600/R-01/020, Newport, OR.

USEPA-USCOE (U.S. Army Corps of Engineers). 1977. Ecological evaluation of proposed discharge of dredged material into ocean waters. Technical committee on criteria for dredged and fill material. Environmental Effects Laboratory. U.S. Army Engineer Waterways Experiment Station. Vicksburg, MS.

USEPA-USCOE (U.S. Army Corps of Engineers). 1991. Evaluation of dredged material proposed for ocean disposal (Testing manual). EPA-68-C8-0 105. USEPA, Office of Marine and Estuarine Protection, Washington, D.C.

USEPA-USCOE (U.S. Army Corps of Engineers). 1994. Evaluation of dredged material proposed for discharge in inland and near coastal waters. EPA 000/0-93/000, Washington, DC.

Virnstein, R.W. 1977. The importance of predation by crabs and fishes on benthic infauna in Chesapeake Bay. *Ecology* 58: 1199-1217.

Walters, D.B. and C.W. Jameson. 1984. *Health and safety for toxicity testing*. Butterworth Publications, Woburn, MA.

Williams, L.R. 1993. Methods for the EPA's regulatory program. *Environmental Testing and Analysis,* 2:36.

Williams, J.B., B.J. Copeland, and R.J. Monroe. 1986. Population dynamics of an r-selected bivalve, *Mulinia lateralis* (Say) in a North Carolina estuary. *Contributions in Marine Science,* 29:73-89.

Williams, L.G., P.M. Chapman, and T.C. Ginn. 1986. A comparative evaluation of marine sediment toxicity using bacterial luminescence, oyster embryo, and amphipod bioassays. *Marine Environmental Research* 19: 225 - 249.

Winer, B.J. 1971. *Statistical Principles in Experimental Design*. 2nd Ed. McGraw-Hill Book Company, New York, NY, 907 p.

Word, J.Q., J.A. Ward, L.M. Franklin, V.I. Cullinan, and S.L. Kiesser. 1987. Evaluation of the equilibrium partitioning theory for estimating the toxicity of the nonpolar organic compound DDT to the sediment dwelling amphipod, *Rhepoxynius abronius.* Battelle Marine Science Research Laboratory Report, Task 1, WA56, Sequim, WA, 60 pp.

Zar, J.H. 1984. Biostatistical Analysis. 2nd Ed. Prentice-Hall, Inc., Englewood Cliffs, NJ, 717 p.

Appendix A

Sample Data Sheets

Mulinia lateralis Juvenile Sublethal Sediment Toxicity Test: Breakdown Sheet

PROJECT: _____ **TEST NUMBER:** _____

DAY: _____ **ANALYST:** _____

DATE: _____

Treatment	Rep	Number Added	Number Found	Number Alive	Mortality (%)	Weight (g)	Mean Weight (g)
Comments:							
Comments:							
Comments:							
Comments:							

***Mulinia lateralis* Juvenile Sublethal Sediment Toxicity Test: Daily Monitoring Data Sheet**

PROJECT: _____ **TEST NUMBER:** _____

DAY: _____ **ANALYST:** _____

DATE: _____

Treatment	Rep	Temperature (°C)	Salinity (‰)	D.O. (mg/L)	pH	Comments

Also Report Water Bath Temperature (°C):

pH Meter Calibration Slope:

Temperature (°C) and Salinity (‰) of Dissolved Oxygen Standard:

Saturated Dissolved Oxygen Concentration (mg/L):

Food Temperature (°C) and Salinity (‰):

Mulinia lateralis Juvenile Sublethal Sediment Toxicity Test: Algal Feeding Sheet

PROJECT: _____ **TEST NUMBER:** _____

DAY: _____ **ANALYST:** _____

DATE: _____

COUNT INFORMATION

COUNTS		MEAN COUNT	HEMOCYTOMETER DILUTION FACTOR	OPERATIONAL DILUTION FACTOR	FINAL CONC. (cells/mL)
Isochrysis					
1	x	1.00 + 04	x	=	
2					
Tetraselmis					
1	x	1.00 + 04	x	=	
2					

CALCULATIONS

Isochrysis	$\frac{\text{Final Conc.}}{5.00E+04}$	=	X	=	$\frac{200\text{ mL}}{X}$	=	I	=	mL of conc. to add
Tetraselmis	$\frac{\text{Final Conc.}}{5.00E+04}$	=	Y	=	$\frac{200\text{ mL}}{Y}$	=	I	=	mL of conc. to add

PREPARATION OF CONCENTRATE

(a) Total volume of algae to be added to each test chamber = I + T =

(b) Species % = mL of concentrate to add/total volume of algae/test chamber (a)

a. *Isochrysis* =

b. *Tetraselmis* =

(c) Total volume of algae to be added to test =

a. total volume of algae/test chamber (a) × # of test chambers =

(d) Amount of each algae in stock = Species % (b) × total volume of algae/test (c)

a. Isochrysis =

b. Tetraselmis =

(e) Add algae together, mix and dispense at volume (a) into each test chamber

COMMENTS: